Longitudinal analysis of mutual fund performance

Ter Horst, J.R.

Publication date:
1998

Document Version
Publisher's PDF, also known as Version of record

CentER
for
Economic Research

Longitudinal Analysis

of Mutual Fund Performance

Jenke R. ter Horst

Tilburg University

Longitudinal Analysis of Mutual Fund Performance

Longitudinal Analysis of Mutual Fund Performance

Proefschrift

ter verkrijging van de graad van doctor aan de Katholieke Universiteit Brabant, op gezag van de rector magnificus, prof. dr. L.F.W. de Klerk, in het openbaar te verdedigen ten overstaan van een door het college van decanen aangewezen commissie in de aula van de Universiteit op

vrijdag 20 november 1998 om 14.15 uur

door

Jenke Rien ter Horst

geboren op 4 oktober 1967 te Den Ham

Promotor: Prof. dr. Th.E. Nijman

Copromotor: Dr. M.J.C.M. Verbeek

Voor mijn ouders

Preface

This study presents the results of the project 'Longitudinal Analysis of Stock Market Returns' financed by the Netherlands Organization for Scientific Research (N.W.O.). Parts of this study were written in cooperation with others: Chapter 3 and 4 is joint work with Theo Nijman and Frans de Roon, Chapter 5 is joint work with Marno Verbeek, and Chapter 6 is joint work with Theo Nijman and Marno Verbeek.

Acknowledgement

This thesis is the result of four years work at the Department of Econometrics at Tilburg University. I will use this opportunity to thank a number of people who created an excellent work environment and who were a great help for me in writing this thesis. First of all, I want to thank my supervisors Theo Nijman and Marno Verbeek. Without Theo and Marno I would never have succeeded in finishing this study within this period. I will not distinguish between them because both deserve many thanks for spending so much time and energy in reading and correcting numerous versions of the underlying chapters of this thesis. They have answered all my questions, stupid or not, and convinced me that I was able to complete the project with a positive result. I will never regret that I applied for this project with Theo and Marno as unknown supervisors for me. I admire them both, also personally, very much appreciate their scientific and personal advises, and hope that we can continue the collaboration in the future.

Second, I would like to thank Frans de Roon. As co-author of two chapters, Frans has delivered a significant contribution to this thesis. He has introduced me in the world of mean-variance spanning, and was always prepared to share new ideas concerning this topic with me. I would also like to thank Angelien Kemna, Piet Moerland, Michael Rockinger, Peter Schotman and Arthur van Soest for their interest in my work and for consenting to participate in my Ph.D. committee.

Furthermore, a special word of thanks also goes to my colleagues and friends at Tilburg University, in particular, to my running, diner and roommates. Besides very much appreciated scientific discussions, they also created a relaxed atmosphere that, in my opinion, is a necessary condition for completion of a Ph.D. project.

Tenslotte wil ik mijn ouders bedanken voor al de steun die zij mij hebben gegeven, en die mij hebben gebracht waar ik nu ben. Het was de laatste vier jaar niet altijd even gemakkelijk voor *mij, maar zij waren een luisterend oor op de momenten dat ik het moeilijk had, en ze hebben op die wijze heel erg veel bijgedragen aan dit proefschrift.*

Jenke ter Horst
August 1998

Chapter 1
Introduction

1.1 Motivation

A large fraction of investments by individuals as well as by institutions takes the form of investments in mutual funds. Many different motivations for investing in mutual funds have been provided in the literature, including the claim that managers of mutual funds have superior knowledge and skills that can be used to outperform the market. Testing of the validity of these claims is complicated by two facts. First, the expected returns on mutual funds show cross-sectional as well as time series variation. Second, mutual funds that did not do very well in the past tend to stop trading, e.g. because of a merger with another fund, more often than other funds. The latter implies that an analysis of returns on mutual funds that are currently traded is possibly affected by so-called survivorship bias. The aim of this thesis is to use longitudinal econometric techniques to test the validity of some of the motivations for investing in mutual funds that have been given in the literature.

A large part of the literature on performance evaluation concentrates on the claim that fund managers have special abilities in selecting stocks that make the fund they manage an interesting investment product with return characteristics that were not attainable for investors before. In order to examine this claim an asset pricing model is specified. The role of the pricing model is that it specifies a normal expected return given the risk profile, and in that way it can be investigated whether a mutual fund realizes an expected abnormal return in addition to the normal expected return. Funds that realize such an additional expected return are referred to as having superior performance or as funds that 'beat the market'.

One of the assumptions that are often made in performance evaluation is that expected returns are time-invariant. The validity of this assumption has implications for the way in which one would like to test the validity of the arguments put forward in favor of investing in mutual funds. Recent papers of for instance Ferson and Schadt [1996] advocate conditional evaluation techniques, taking the predictability of returns explicitly into account. Conditional performance evaluation is motivated by the fact that the use of dynamic strategies that are based on publicly available information should not be interpreted as delivering abnormal return.

If expected (abnormal) mutual fund returns are time-varying, the weights in an optimal investment strategy in these funds will usually also be time-varying since one will put more weight on the fund that currently has a higher expected (abnormal) return. Such a dynamic investment strategy can significantly increase the expected return on a portfolio of mutual funds if fund returns are predictable. The hypothesis that expected (abnormal) returns on mutual funds vary over time is known as the hypothesis of persistence in performance of the funds. This hypothesis also receives a lot of attention in the empirical finance literature (see, e.g. Carhart [1997a]). Mutual fund managers that are characterized by positive autocorrelation in returns and recently showed a good performance are referred to as having 'hot hands' (see, e.g. Hendricks, Patel and Zeckhauser [1993]). A recent paper of Sirri and Tufano [1998] indicates that individual investors behave as if a persistence pattern is present, investing more in funds that recently showed a good performance.

As stated, performance analysis of mutual funds is plagued by the fact that less successful mutual funds tend to stop trading more easily. As argued by Brown, Goetzmann and Ross [1995] and Hendricks, Patel and Zeckhauser [1997], performance analyses based on currently traded funds that do not take this phenomenon into account can easily generate misleading inference due to this so-called survivorship bias.

1.2 Overview and Contribution of the Thesis

Chapter 2 of this thesis presents an extensive overview of empirical mutual fund performance evaluation studies, together with the main arguments that have been proposed in the literature for investing through mutual funds. An important argument that will be addressed is that mutual funds can extend the investment set by reducing the required risk to reach a prespecified expected portfolio return. The chapter discusses a number of empirical studies that examine whether mutual funds that recently showed good performance are likely to continue good performance in the near future. As mentioned above, the methods that are used to measure mutual fund performance and its persistence are possibly affected by survivorship effects. Chapter 2 also gives a short review on the literature on survivorship bias and discusses the empirical findings on the size of the bias on e.g. expected (abnormal) returns and persistence measures. Moreover, a number of possible approaches will be analyzed that have been proposed in the literature to handle the problem of survivorship bias.

Chapter 3 discusses how regression analysis can be used to test the claim that mutual funds extend the investment set of the investor's current portfolio. The testing procedure that is proposed is a direct extension of the well known performance measure proposed by Jensen [1968] and is based on the recent literature on testing for extensions of the investment set (see, e.g. Huberman and Kandel [1987], Bekeart and Urias [1996] and DeRoon, Nijman and Werker [1996]). Since investors are confronted with short selling restrictions and transaction costs, an important shortcoming of the standard tests for extensions of the investment set is the underlying assumption of frictionless markets. Recent papers of Luttmer [1996] and DeRoon, Nijman and Werker [1998] have developed tests that take these market frictions into account. Chapter 3 empirically analyzes the performance of internationally investing U.S. based mutual funds in a frictionless market as well as a market with short sales constraints and transaction costs. Moreover, we test for outperformance of the mutual funds in an unconditional as well as a conditional framework.

Most mutual fund performance evaluation studies start with comparing realized mutual fund returns with a benchmark asset that corresponds to the self-reported investment style of the mutual fund. The disadvantage of this so-called relative performance evaluation is that the actual investment style does not always correspond to the reported style, as shown by Brown and Goetzmann [1997]. Chapter 4 of this thesis shows how return-based style analysis, introduced by Sharpe [1992], can be used to avoid the self-reported investment styles. Moreover, the chapter shows how style analysis can be used as a valuable instrument in relative performance evaluation in order to determine the exposure to asset classes. Subsequently, relative performance evaluation is related to testing whether mutual funds extend the investment set of the investor's current portfolio. Empirical estimates of the performance of a sample of Dutch mutual funds, both in terms of performance relative to the reported style and to the estimated style as well as in terms of measures like Jensen's alpha are also presented in Chapter 4.

A number of papers (see e.g. Jegadeesh [1990], Hendricks, Patel and Zeckhauser [1993]) have addressed the estimation and testing of persistence in performance if only data for a small number of time periods is available or if one is only willing to assume that the parameters in the model are constant for a small number of periods. Several proposed estimation procedures have a bias that is inversely related to the number of periods T in the sample. Chapter 5 analyzes how this bias depends on the parameters in the model and discusses to what extent recent findings in favor of a 'hot hand' phenomenon can be explained by this bias. Analytical expressions, combined with a simulation study, show that the importance of these biases cannot be neglected for the samples that are typically used in applied work, in particular if the number of time periods is small. Furthermore, we propose an instrumental variables estimator of the persistence parameters that does not suffer from biases in the order of T^{-1} and present empirical estimates of the size of the persistence effect for U.S. based mutual funds that predominantly invest in U.S. growth stocks.

Chapter 6 of this thesis considers the estimation of persistence in performance in case survival depends on past returns. We specify a longitudinal probit model for the probability of survival of a mutual fund. Obviously the past record of fund returns will be one of the most important variables that determine the probability of survival besides the time since fund inception. The probit model that is proposed is an extension of the model of Brown and Goetzmann [1995] by allowing for aggregate macro-economic shocks. In a number of simulation experiments, where we apply the probit specification to determine the probability of survival, it appears that for instance the expected returns of samples of surviving funds are upward biased. Moreover, a spurious persistence in performance pattern is found in a sample plagued by survivorship. After having obtained knowledge and insight in the size of survivorship effects on various processes, we propose a fairly simple weighting method based on the use of estimated survival probabilities that eliminates the survivorship biases that arise using traditional techniques.

Chapter 2

A Survey of the Literature on Mutual Fund Performance Evaluation

2.1 Introduction

The mutual fund industry has grown considerably. Since the early 1920's when the first mutual funds were formed in the United States, the amount invested in mutual funds at the end of 1996 has grown to more than a trillion dollars in the U.S. and to more than 100 billion guilders in the Netherlands. In the early days all mutual funds had similar investment policies, mainly offering a diversified portfolio of common stocks and bonds. Nowadays, there is an enormous diversification in the types of mutual funds, varying from investment policies within particular industry groups to investing mainly in emerging markets. Together with the growth in the number of funds, the number of papers published on evaluating mutual fund performances and its persistence has grown significantly.

One of the main motives for the research in these papers is to examine why investors invest in mutual funds. The following reasons may explain the observed behavior of investors (see, e.g. Gruber [1996]):

1. Mutual funds have a professional management that is able to select the right stocks at the right moment.
2. Mutual funds are an easy way to obtain a diversified portfolio of stocks.
3. Mutual funds have low transaction costs compared with buying individual stocks.
4. The investment companies managing various mutual funds offer the opportunity to switch among funds within the company without much effort.

Although the second, third and fourth argument can be important motivations to invest in mutual funds, the first motivation gets most attention in mutual fund evaluation studies. In Section 2.2 of this survey we will present a number of performance evaluation techniques that are used to analyze the validity of this first argument. This will be illustrated with some empirical results from recent performance evaluation studies.

The realization of high average past returns relative to other mutual funds or market indices appears to be an important motive to invest in a mutual fund, as stressed by, for instance, Rockinger [1995] and Sirri and Tufano [1997]. In order to examine whether investors behave irrationally in choosing funds with a currently good performance, the persistence in performance of mutual funds gets a lot of attention in the performance evaluation literature. In Section 2.3 of this survey we will present a number of methods that are used to examine whether funds with 'good' performance in the recent past are more likely to exhibit 'good' performance in the future, i.e. whether 'good' performance is persistent. This will be illustrated with some empirical estimates of persistence from recent performance studies.

The starting point in performance evaluation studies is the construction of the sample of funds to be analyzed. Commercially available datasets are characterized by the fact that they only contain mutual funds that are still in existence. Early studies on performance evaluation (e.g. Jensen [1968]) did not take notice of the potential effect of observing only the surviving mutual funds. Ignoring this effect creates the possibility of significant biases in performance evaluation studies (see, e.g. Grinblatt and Titman [1989a]). For instance, if mutual funds with high past returns have a higher probability to survive than mutual funds with low past returns, the average return of the sample of mutual funds will be overestimated. Recent studies on performance evaluation (e.g. Malkiel [1995], Carhart [1997a]) claim that their analyses do not suffer from survivorship effects anymore. In these studies, the data on all the funds that did not survive until the end of the sample period is also taken into account until the moment of fund disappearance. However, Carhart [1997b] identifies a so-called look ahead bias that might still be present in some methods. In Section 2.4 of this survey we will discuss the problem of survivorship bias and look ahead bias in more detail. Methods will be presented that have been proposed to measure the size of the potential effect on various performance measures. These will once more be illustrated with some empirical evidence from the recent literature.

2.2 Measuring Mutual Fund Performances

2.2.1 Introduction

One of the possible explanations for the enormous popularity of mutual funds is that the managers of these mutual funds have special abilities which provide individual investors with superior returns that could not have been attained by the investors themselves. To determine whether fund managers offer an investment product with certain characteristics that was not attainable for individual investors before the fund was taken into account, is an important question that performance evaluation tries to answer.

A simple measure of performance is to compare the average return of a mutual fund with the average returns on other mutual funds or market indices like, for instance, the S&P 500. Although this so-called relative performance evaluation is often applied, it suffers from at least two potential problems. A first problem is that according to standard portfolio theory an asset with a high expected return is not necessarily a more attractive investment than an asset with a lower expected return. A second problem is the choice of an appropriate reference group or benchmark portfolio (see, e.g. Wermers [1997]). Most often, average mutual fund returns are compared with those mutual funds with a similar investment objective or a benchmark asset reflecting the fund's investment objective. However, this method can easily be gamed by the fund managers since mutual fund managers have to report the investment style themselves (see, e.g. Brown and Goetzmann [1997]). Consequently, the actual investment style does not necessarily correspond to the reported style, and a wrong reference group or benchmark asset could be used in performance evaluation. Moreover, for instance, equity funds do not invest all wealth in stocks but hold some cash as well. Style analysis can be used to obtain an objective estimate for the fund's investment style and the allocation to certain asset classes (see, e.g. Sharpe [1992]). This will be extensively discussed in Section 2.2.6 and Chapter 4 of this thesis.

The remainder of this section is organized as follows. In Section 2.2.2 we will discuss performance evaluation methods where the portfolio that an investor initially holds is taken into account. We will illustrate this in Section 2.2.3 with some empirical examples of recent performance evaluation studies. In the Sections 2.2.2 and 2.2.3 it is assumed that expected returns and (co)variances are constant over time. This assumption will be relaxed in the Sections 2.2.4 and 2.2.5 where the so-called conditional performance evaluation techniques will be discussed and illustrated. In Section 2.2.6 we will return to constant expected returns and (co) variances and discuss style analysis and the link between style analysis and performance evaluation in this setting. Recently, new data sets have become available that also contain information about mutual fund portfolio compositions. In Section 2.2.7 we will refer to some performance evaluation techniques that explicitly take this additional information into account.

2.2.2 Performance Measurement in a Portfolio Context

In this section we discuss several models that form the basis of modern performance measurement and outline the underlying assumptions. Consider an individual investor that currently invests in K assets with return vector R_{t+1}. The K-dimensional vector of expected returns is denoted as

$$\mu_R = E[R_{t+1}] \tag{2.1}$$

with corresponding covariance matrix

$$\Sigma_{RR} = V[R_{t+1}]. \tag{2.2}$$

Suppose the investor considers to extend his initial efficient set of K assets by adding a set of N mutual funds. The return on this set of mutual funds is denoted as r_{t+1}, while the corresponding expected return vector and covariance matrix are denoted as

$$\mu_r = E[r_{t+1}] \tag{2.3}$$

and

$$\Sigma_{rr} = V[r_{t+1}]. \tag{2.4}$$

In order to judge whether this set of N funds extends the mean-variance efficient set, the covariance with the initial set of K assets has to be taken into account. Let this covariance be denoted as Σ_{rR}. The vector of initial portfolio weights is referred to as w_R, while the extended weight vector is denoted as w. In the sequel, we refer to the extended set when a subscript is absent in the notation. Moreover, we assume, unless explicitly stated differently, a frictionless market in evaluating mutual fund performances, i.e. no transaction costs and short selling of assets is allowed in unrestricted amounts.

For a mean-variance investor, the optimal weight vector for the initial K assets can be written as (see Appendix 2.A or any standard textbook in finance for the derivation):

$$\tilde{w}_R = \tilde{\gamma}^{-1}\Sigma_{RR}^{-1}(\mu_R - \eta\iota_K), \tag{2.5}$$

where $\tilde{\gamma}$ is the investor's risk aversion coefficient, ι_K is a K-vector of ones and η is the expected return on the zero beta portfolio of $\tilde{w}_R$, which can be obtained as the intercept of the line tangent to the mean-variance frontier at $\tilde{w}_R$. From Appendix 2.A it also follows that the zero beta rate η depends on the risk aversion coefficient $\tilde{\gamma}$. If the investor cannot extend the investment set by investing in the set of N mutual funds, the extended optimal weight vector will have the following form $w = (\tilde{w}_R\ , 0_N)'$. It is straightforward to show that if this portfolio choice is efficient, the relationship

$$\mu - \eta\iota_{K+N} = \tilde{\gamma}\begin{pmatrix} \Sigma_{RR} & \Sigma_{Rr} \\ \Sigma_{rR} & \Sigma_{rr} \end{pmatrix}\begin{pmatrix} \tilde{w}_R \\ 0_N \end{pmatrix} \tag{2.6}$$

will hold for the extended set of $K + N$ assets. It is obvious that the first K elements of (2.6), i.e. the optimal weights for the initial K assets, coincide with the expression given in (2.5). Moreover, substituting (2.5) into (2.6) gives for the last N rows of (2.6)

$$\mu_r - \eta\iota_N = B(\mu_R - \eta\iota_K), \tag{2.7}$$

where $B \equiv \Sigma_{rR}\Sigma_{RR}^{-1}$ is of dimension $N \times K$. Recall that η is the zero beta rate corresponding to investor's initial portfolio of K assets. It is important now to distinguish two different cases. First of all it is possible that (2.7) only holds for one value η, this can be interpreted as stating that the mean-variance efficient portfolio that the investor was holding, is also mean-variance efficient on the extended set of $K + N$ assets. Consequently, the two mean-variance frontiers will intersect at the investor's initial portfolio location.

Next to the possibility that there is only one value of the risk aversion coefficient for which the investor cannot extend the investment set, there is the possibility that the investor cannot extend the mean-variance efficient set by taking a position in the set of N mutual funds independent of the risk aversion coefficient. If this is the case then (2.7) should hold for all $\tilde{\gamma}$ (and therefore for all η), implying that

$$\iota_N - B\iota_K = 0 \text{ and } \mu_r = B\mu_R. \tag{2.8}$$

Following the same reasoning as above, (2.8) can be interpreted as stating that the mean-variance frontier of the K plus the N funds will coincide with the frontier of the initial K assets.

As shown by, for instance, Jobson and Korkie [1989], the question whether investors can extend the investment set by investing in a set of mutual funds is closely related to performance measurement. In order to evaluate the performance of a mutual fund and to define outperformance a pricing model is required that specifies the set of K efficient benchmark portfolios that span the mean-variance frontier. This implies that performance measurement and looking for an extension of the efficient set are equivalent under the assumption that the K initial assets the investor holds are already an efficient combination of the benchmark assets corresponding to the pricing model.

The vector

$$\alpha_J(\eta) \equiv \mu_r - \eta\iota_N - B(\mu_R - \eta\iota_K), \tag{2.9}$$

is known as the vector of generalized Jensen measures. A positive element in the vector of generalized Jensen measures $\alpha_J(\eta)$ indicates outperformance of the benchmark assets by the corresponding mutual fund, while a negative element is interpreted as underperformance. Moreover, it is straightforward to show that, at least in the scalar case where $N = 1$, an investor can extend the investment set by investing in a mutual fund that shows outperformance by taking a long position in the mutual fund under consideration (see equation (2.16) below and Appendix 2.A for details). A similar analysis shows that a negative value for $\alpha_J(\eta)$ implies a short position in the mutual fund for the efficient portfolio.

The vector $\alpha_J(\eta)$ generalizes the original alpha-measure proposed by Jensen [1968] in at least four ways. First of all Jensen assumed that one of the benchmark assets is a risk free deposit. In that case the zero beta rate will be equal to the risk free rate. Secondly, Jensen [1968] assumed the validity of a specific pricing model. If a valid pricing model implies that the K benchmark portfolios span the efficient frontier, based on public information, positive values for Jensen's alpha can be attributed to superior investment skills. Thirdly, Jensen considered the case where only two assets (the risk free deposit and the market portfolio in his case) span the frontier, while K in (2.9) is not restricted to be 2. Finally, (2.9) considers the simultaneous addition of N mutual funds to the initial portfolio, rather than just one as in Jensen [1968].

As shown by Huberman and Kandel [1987], the hypothesis that there is only one value of the risk aversion coefficient for which the investor cannot extend the investment set by investing in the set of mutual funds, i.e. $\alpha_J(\eta) = 0$ for one η, or the hypothesis that this holds for all values of the risk aversion, i.e., that the restrictions in (2.8) are satisfied, can easily be tested in a regression framework. Substituting realized returns r_{t+1} and R_{t+1} for the expected returns μ_r and μ_R in (2.7) gives

$$r_{t+1} = \alpha + BR_{t+1} + \varepsilon_{t+1}, \tag{2.10}$$

where $\alpha = \mu_r - B\mu_R$ and the idiosyncratic error term

$$\varepsilon_{t+1} = (r_{t+1} - \mu_r) - B(R_{t+1} - \mu_R) \tag{2.11}$$

is by definition uncorrelated with the return on the K initial assets and has expectation zero. The properties of the error term ε_{t+1} imply that equation (2.10) can be consistently estimated by Ordinary Least Squares.

The hypothesis that only for the particular value of the risk aversion coefficient corresponding to the zero beta rate η the investor cannot extend the efficient set by including a set of mutual funds can now be framed as

$$H_0 : \alpha - (\iota_N - B\iota_K)\eta = 0, \tag{2.12}$$

while the hypothesis that this holds for all possible η is given by

$$H_0 : \alpha = 0 \text{ and } \iota_N - B\iota_K = 0. \tag{2.13}$$

Both hypotheses can be tested using a standard Wald test, which, under the null-hypothesis, is χ^2 distributed with N and $2N$ degrees of freedom, respectively (see, e.g. Huberman and Kandel [1987]). Rejection of the hypothesis implies a significant extension of the efficient investment set. Alternatively, one can say that rejection implies an abnormal performance of the mutual funds under consideration with respect to the K benchmark assets, where abnormal can imply out as well as underperformance of the K initial assets. Note that the left hand side of (2.12) equals the generalized Jensen measure $\alpha_J(\eta)$. Consequently, testing for abnormal performance in a particular point is equivalent to testing whether the generalized Jensen measure is equal to zero for the corresponding zero beta rate.

As an aside we now introduce another performance measure that is often used in performance evaluation studies. The so-called Sharpe ratio (Sharpe [1966]), defined as:

$$\theta(\eta) = \frac{w_R' \mu_R - \eta}{\sqrt{w_R' \Sigma_{RR} w_R}}, \tag{2.14}$$

measures the expected excess return on a portfolio per unit of risk exposure. Note that there is an important difference with the Jensen measure. The Sharpe measure only takes into account the characteristics of one portfolio, i.e. the expected return and the corresponding standard deviation, while the Jensen measure explicitly takes the covariance of a portfolio with the initial set of assets into account. Therefore, these two performance measures answer different questions. The Sharpe measures answers the question whether portfolio A should be preferred over portfolio B or vice versa, in case an individual investor is restricted to invest in either portfolio A and a riskless deposit with return η or in portfolio B and the riskless deposit. The Jensen measure answers the question whether investors can extend the efficient set by investing in portfolio A, B or both, given that the individual investor already holds an efficient portfolio of K assets.

In Appendix 2.A it is shown that the following relationship between the squared Sharpe measure of efficient portfolios and the Jensen measure holds:

$$\theta_{N+K}^2(\eta) = \theta_K^2(\eta) + \alpha_J(\eta)' \Sigma_{\varepsilon\varepsilon}^{-1} \alpha_J(\eta), \tag{2.15}$$

where $\theta_{N+K}^2(\eta)$ and $\theta_K^2(\eta)$ are the squared Sharpe measures of respectively the extended set of $N + K$ assets and the initial set of K assets, and where the generalized Jensen measure $\alpha_J(\eta)$ and the inverse of the covariance matrix of ε_{t+1}, i.e. $\Sigma_{\varepsilon\varepsilon}^{-1}$, can both be obtained from the regression (2.10).

The coefficients α and B in (2.10) cannot only be used to test (2.12) and (2.13), but can also be used to construct the new optimal weights of the extended set of assets given knowledge of the initial optimal weights $\tilde{w}_R$. In Appendix 2.A it is shown that for a given value η, the new optimal portfolio weights can be written as

$$w_r = \left(\frac{\tilde{m}-\eta}{\theta_K^2(\eta)+(\tilde{m}-\eta)\alpha_J(\eta)'\Sigma_{\varepsilon\varepsilon}^{-1}(\iota_N - B\iota_K)}\right)\Sigma_{\varepsilon\varepsilon}^{-1}(\alpha_J(\eta)) \tag{2.16}$$

and

$$w_R = \left(\frac{\theta_K^2(\eta)}{\theta_K^2(\eta)+(\tilde{m}-\eta)\alpha_J(\eta)'\Sigma_{\varepsilon\varepsilon}^{-1}(\iota_N - B\iota_K)}\right)\tilde{w}_R - B'w_r, \tag{2.17}$$

where as before $B \equiv \Sigma_{rR}\Sigma_{RR}^{-1}$, $\Sigma_{\varepsilon\varepsilon}^{-1} \equiv (\Sigma_{rr}-\Sigma_{rR}\Sigma_{RR}^{-1}\Sigma_{Rr})^{-1}$ and $\theta_K^2(\eta)$ is the squared Sharpe measure of efficient portfolios in the initial investment problem with expected return $\tilde{m}$. Note that B is equal to the matrix of coefficients of R_{t+1} that can be obtained in a regression of r_{t+1} on a constant and R_{t+1}, i.e. regression equation (2.10). Consequently, the optimal weights in the mutual funds, i.e. (2.16) can be completely obtained from a regression of r_{t+1} on a constant and R_{t+1} and knowledge of the initial efficient portfolio $\tilde{w}_R$. Note that in (2.16) and (2.17) the zero beta rate η is fixed at the rate that corresponds to the initial efficient portfolio. Consequently, the new optimal portfolio will have an expected return equal to m $(= w'\mu)$, while the old optimal portfolio has an expected return $\tilde{m}$ $(= \tilde{w}_R'\mu_R)$. Under the assumption that a risk free asset is available, DeRoon [1997] gives expressions for the new optimal weights under the condition that $m = \tilde{m}$.

From equation (2.16) it follows that in case the additional set of N mutual funds that the investor is taking into account contains only one mutual fund, i.e. $N = 1$, the sign of $\alpha_J(\eta)$ determines whether the investor should take a long or short position in the additional asset. In case of $N > 1$, the inverse of the covariance matrix of ε_{t+1}, $\Sigma_{\varepsilon\varepsilon}^{-1}$, as well as the sign of $\alpha_J(\eta)$ determine the sign of the weight in the additional assets in the extended investment problem. From (2.15) it follows that the Sharpe ratios of the extended set of assets and the initial set of assets are identical in case of a generalized Jensen measure equal to zero. Consequently, the investor cannot improve the risk-return trade-off of his initial portfolio by investing in the additional set of mutual funds.

Thusfar, we avoided assumptions on the validity of any pricing model. In contrast, a large part of the performance evaluation literature starts with assuming a pricing model. The models that are used vary between single factor models such as the traditional CAPM and multi factor extensions such as the Arbitrage Pricing Theory (APT). Widely known anomalies of the single index model, such as the failure to adequately explain the cross-sectional variation in expected returns between firms with a small or large market capitalization, the so-called size effect, have lead to the development of multi factor models. Recall that testing whether an investor can extend the investment set by investing in a set of mutual funds and outperformance by mutual funds are identical under the assumption that investors hold efficient combinations of the K benchmark assets that correspond to the pricing model used. Moreover, the benchmark assets can be interpreted as factor mimicking portfolios for the K risk factors of the supposed pricing model. Accordingly, the estimate for the generalized Jensen measure can also be interpreted as a risk-adjusted return.

Finally, it has to be mentioned that a common assumption in the performance evaluation literature is the absence of market frictions. However, when investing in assets or mutual funds, individual investors are confronted with transaction costs and short sales restrictions. Moreover, mutual funds have operating expenses such as management fees, administrative costs, advisory fees and marketing costs which are deducted from the fund's asset value. In performance measurement these expenses are often taken into account by distinguishing between mutual fund returns before expenses and after expenses are subtracted (see, e.g. Malkiel [1995]). However, the load-fees charged by some of the mutual funds are usually ignored. In Chapter 3 of this thesis, we evaluate the performances of mutual funds where we explicitly take into account the short sales restrictions and the transaction costs an individual investor is confronted with.

2.2.3 Evaluating Performances of Unconditional Strategies

In order to discuss whether the argument that mutual funds have a professional management that is able to select the right stocks at the right moment is a valid motive for individual investors to invest in mutual funds, we present some empirical results of the recent performance evaluation literature. Before discussing these results, we first of all make a distinction between actively and passively managed mutual funds. Actively managed mutual funds try to outperform an initial set of K benchmark assets, whereas passively managed mutual funds only try to replicate benchmark assets. Since having a professional management that tries to outperform K benchmark assets implies high management fees, passive mutual funds usually have much lower operating expenses than actively managed mutual funds. Note that Gruber [1996] reports that, although the expense ratio of passive mutual funds is on average lower than the expense ratio of actively managed funds, there is a lot of variation in the ratios' magnitude, e.g., due to the set of benchmark assets that has to be replicated. The operating expenses, together with administrative costs, advisory fees and marketing costs are subtracted from the fund's assets. Therefore, since open-end mutual funds sell at their net asset value, investors are confronted with these expenses, although not directly.

As discussed in Section 2.2.1, the simplest measure of performance is to compare the average return of a mutual fund with average returns of other mutual funds or with average returns on benchmark assets like the S&P 500. In this so-called relative performance evaluation, it is usually found that U.S. based equity mutual funds show underperformane relative to a value weighted market index. For instance, Gruber [1996] reports an underperformance of 1.94% per year over the period 1985-1994, and Malkiel [1995] an underperformance of 1.83% per year over the period 1982-1991. Wermers [1997] reports a before expenses outperformance of 1.29% per year over the period 1975-1994, but finds a before expenses underperformance of 0.53% per year over the period 1983-1994. To illustrate that the choice of the reference portfolio or benchmark is important, Wermers also compares the mutual fund returns with the return on an equally weighted market index. In that case the mutual funds heavily underperform the market index with 3.46% per year. Since during this sample period small stocks realized much higher returns than large stocks, the observed outperformance of mutual funds with respect to the value weighted index is probably due to the fact that mutual funds more heavily invested in small stocks than the corresponding weight in the value weighted market portfolio (see also Brown and Goetzmann [1995]).

Recall from Section 2.2.2 that a mutual fund with the highest average return within a group of funds is not necessarily an attractive investment. In order to judge whether it is optimal for a mean-variance investor to invest a fraction of his wealth in a given mutual fund, the covariance of the fund with the initial assets in his portfolio has to be taken into account. The recent literature on modelling the cross-sectional variation of stock returns (see, e.g. Fama and French [1996] and Chan, Jegadeesh and Lakonishok [1996]) finds that many significant explanatory variables can be found, like for instance, size, book-to-market ratio, earnings-to-price ratio and cash flow-to-price ratio. The single index model fails to adequately explain a lot of the observed cross-sectional variation in stock returns. Nevertheless, most mutual fund performance evaluation studies report Jensen measures under the assumption that the traditional CAPM is the appropriate pricing model. These results can be interpreted as assuming that the market portfolio and the risk free asset are the initial assets of the individual investors.

More formally, in order to evaluate the performance of mutual funds with a one-factor model, the following regression equation is estimated:

$$r_{i,t+1} - r_{f,t+1} = \alpha_i + \beta_i(r^m_{t+1} - r_{f,t+1}) + \varepsilon_{i,t+1}, \tag{2.18}$$

where $r_{i,t+1}$ is the return on mutual fund i in period $t+1$, r^m_{t+1} is the corresponding return on the market portfolio and $r_{f,t+1}$ is the return on a risk free asset. In this model, a test for abnormal performance is equivalent to testing whether $\alpha_i = 0$. Notice that α_i in (2.18) is the original Jensen's alpha as introduced by Jensen [1968]. Recall that a positive value for α_i indicates that the individual investor that is currently investing in the market portfolio and the risk free asset, can extend the efficient set by taking a long position in the mutual fund under consideration, while a negative value indicates a short position in the fund. Alternatively, a test for $\alpha_i = 0$ can be interpreted as a test for the validity of the CAPM as the right pricing model.

Using (2.18) with the S&P 500 index as the market portfolio, Malkiel [1995] reports an average yearly underperformance of 3.20% over the period 1982-1991. Moreover, Malkiel finds negative estimates for the Jensen measure for most of the funds, 19 mutual funds out of the total sample of 239 funds have a significant negative estimate for α_i. Malkiel's findings are confirmed by Gruber [1996], who reports an average yearly market-risk adjusted underperformance of 1.56% over the period 1985-1994. This means that investors currently investing in assets that are reflected by this market index can extend the efficient set by taking a short position in most of the mutual funds under consideration.

As mentioned before, in measuring mutual fund performance one has to note that mutual funds have operating expenses which are subtracted from the fund's asset value. Therefore, it can be the case that mutual funds have special abilities, but, if the expenses are too high, the ability is no longer visible in evaluating returns after expenses. Malkiel [1995] also evaluates mutual fund returns before expenses are subtracted, but still finds an average Jensen measure of -2.03%. In contrast, Daniel, Grinblatt, Titman and Wermers [1997] report a before expenses outperformance of 0.60% over the period 1975-1994. However, these results are not really comparable because of the different market portfolios used in the analyses. Daniel et al. use the much broader CRSP value weighted market index. Moreover, when Daniel et al. correct their reported outperformance for the average expenses, an underperformance is found. So, although the mutual funds seem to have outperformance for individual investors whose initial portfolio is covered by the CRSP index, efficient portfolios still contain a short position in the mutual funds due to the operating expenses that have to be taken into account.

In order to answer the question whether individual investors whose current portfolio is reflected by a market, a size, a book-to-market and a momentum index can extend the efficient set by investing in mutual funds, Carhart [1997a] proposed to test if $\alpha_i = 0$ in a four-factor extension of (2.18), i.e.

$$r_{i,t+1} - r_{f,t+1} = \alpha_i + \beta_{mi}(r^m_{t+1} - r_{f,t+1}) + \beta_{si} r^{smb}_{t+1} + \beta_{hi} r^{hml}_{t+1} + \beta_{pi} r^{pr1yr}_{t+1} + \varepsilon_{i,t+1}, \quad (2.19)$$

where r_{t+1}^{smb} is the difference between the return on a portfolio of small stocks and a portfolio of big stocks, r_{t+1}^{hml} is the difference between the return on a portfolio of high book-to-market and a portfolio of low book-to-market stocks and r_{t+1}^{pr1yr} is the difference between the return on a portfolio of stocks with the highest return over the previous year and a portfolio of stocks with the lowest return over the previous year. Notice that because of the assumption of the existence of a risk free asset in (2.18) as well as in (2.19), that a test for $\alpha_i = 0$ is equivalent to testing whether the mean-variance frontier of the extended set of assets coincides with the mean-variance frontier of the initial assets in portfolio. Alternatively, equation (2.19) can be interpreted as a pricing model with four risk factors, where r_{t+1}^{m}, r_{t+1}^{smb}, r_{t+1}^{hml} and r_{t+1}^{pr1yr} reflect the factor mimicking portfolios.

In contrast to Gruber's [1996] four-factor model, Carhart [1997a] uses a one-year momentum in stock returns portfolio instead of a bond index as fourth factor. The use of the one-year momentum portfolio is based on the failure of Fama and French's [1993] three-factor model to explain cross-sectional variation in momentum sorted portfolio returns (see Jegadeesh and Titman [1993] and Chan, Jegadeesh and Lakonishok [1996]). Notice that the models mentioned give in fact answers to different questions. For instance, the three-factor model gives answer to the question whether an investor that currently follows a strategy of investing in a market portfolio, a size portfolio and a book-to-market portfolio can extend the efficient set by investing in a set of mutual funds. This in contrast to the four-factor model, that also supposes that the investor already follows a momentum strategy. Consequently, the mentioned anomaly of the three-factor model can also be interpreted to imply that investors who do not follow a momentum strategy yet, can extend the efficient set by taking a position in a mutual fund that follows such a momentum strategy.

Unfortunately, Carhart does not report average Jensen measures based on his four-factor model. Carhart concentrates on the predictability of these Jensen measures, a topic that will be discussed in Section 2.3 of this survey and Chapter 5 and 6 of this thesis. Daniel et al. [1997] apply Carhart's four-factor model on mutual fund returns before expenses. They report an average yearly outperformance of 0.39% over the period 1975-1994. Moreover, Daniel et al. also examine the performance of funds with the same investment objective. It appears that funds with the investment objective 'balanced and income' show a before expenses outperformance of 0.82% per year. The average expenses of the mutual funds are on average almost 1.00% percent, indicating that a short position in most of the funds provides the individual investor an extension of his efficient portfolio. Finally, note that funds that charge a load fee almost always underperform no load funds (see, Gruber [1996]).

2.2.4 Conditional Performance Evaluation and Timing Strategies

Thusfar we assumed that expected returns and (co)variances are constant over time. However, in recent studies evidence is found that stock and bond returns are predictable over time (see, e.g. Ferson and Harvey [1993], Keim and Stambaugh [1986]). Dividend yields, interest rates or some other variables appear to be valuable instruments in determining expected stock returns. If expected returns are time-varying, the weights in an optimal investment strategy will also be time-varying. Consequently, mean-variance optimizing investors will dynamically adjust their portfolios because of changing economic conditions. Interpreting this in the framework sketched in Section 2.2.2, this implies that under certain economic circumstances an investor can improve the risk-return trade-off of his initial portfolio by investing in a set of N mutual funds while under other circumstances the diversification benefits are absent. A complicating factor in conditional performance evaluation is that the optimal weight vector (and thus the mean-variance frontier) changes over time.

Let us denote z_t as the L-dimensional vector of information variables supposed to reflect the state of the economy, and define Z_t as $Z_t \equiv (1\ z_t')'$. Assume that the K-dimensional vector of expected returns on the initial set of assets is denoted as

$$E[R_{t+1}|Z_t] = \mu_R^z. \tag{2.20}$$

with corresponding conditional covariance matrix

$$V[R_{t+1}|Z_t] = \Omega_{RR}^z. \tag{2.21}$$

The conditional expected return and covariance matrix of the additional set of N mutual funds are defined in a similar way, but with subscript r. Equivalent to the unconditional case, in order to judge whether the set of N mutual funds provides an extension of the efficient set, the conditional covariance of the initial set of K assets with the set of N mutual funds has to be taken into account. Let this covariance be denoted as

$$Cov[r_{t+1}, R_{t+1}|z_t] = \Omega_{rR}^z. \tag{2.22}$$

In case that the investor cannot extend the efficient set by investing in the set of N mutual funds, it is straightforward to show that for a given value of the zero beta rate η, the last N elements of the extended $(N+K)$-dimensional optimal weight vector w can be written as:

$$\mu_r^z - \eta\iota_N = B^z(\mu_R^z - \eta\iota_K), \tag{2.23}$$

where $B^z \equiv \Omega_{rR}^z(\Omega_{RR}^z)^{-1}$. Note that this is the conditional analogue of (2.7).

A number of alternative ways are proposed in the literature to incorporate conditional information in mutual fund performance evaluation. Ferson and Schadt [1996] for instance, assume that each of the elements in B^z is linear in the set of variables reflecting the state of the economy, Z_t. Consider the case of one additional mutual fund, i.e. $N = 1$. Following Shanken [1990] and others, they approximate the elements in B^z linearly, i.e.

$$B_k^z = b_{0k} + b_k' z_t, \tag{2.24}$$

where B_k^z denotes the k-th element of the $1 \times K$ dimensional row vector B^z, and where $b_k' \equiv (\tilde{b}_{1k}...\tilde{b}_{lk})$ is a $1 \times L$ row vector. Denote $b_0' \equiv (b_{01}...b_{0k})$ and $b^z \equiv (b_1' z_t...b_k' z_t)$. Substituting realized returns for the conditional expected returns in (2.23) gives

$$r_{t+1} = \alpha + b_0' R_{t+1} + b^z R_{t+1} + \varepsilon_{t+1}, \tag{2.25}$$

where $\alpha = \mu_r^z - B^z \mu_R^z$, where

$$\varepsilon_{t+1} = (r_{t+1} - \mu_r^z) - B^z(R_{t+1} - \mu_R^z). \tag{2.26}$$

Using the same framework as in Section 2.2.2, this implies that for a given z_t, say $\tilde{z}$, the hypothesis that only for a particular value of the risk aversion coefficient the investor cannot extend his efficient set by investing in the set of mutual funds, is equivalent to testing whether

$$\alpha - \eta(1 - b_0' \iota_K - b^{\tilde{z}} \iota_K) = 0, \tag{2.27}$$

while the hypothesis that this holds for all possible risk aversion coefficients is equivalent to

$$\alpha = 0 \text{ and } 1 - b_0' \iota_K - b^{\tilde{z}} \iota_K = 0. \tag{2.28}$$

For arbitrary values of the set of economic variables, these hypotheses can be framed as

$$\alpha - \eta(1 - b_0' \iota_K) = 0 \text{ and } b_k = 0 \tag{2.29}$$

for a particular value of the risk aversion coefficient and

$$\begin{aligned} \alpha &= 0 \\ 1 - b_0' \iota_K &= 0 \\ b_k &= 0 \end{aligned} \tag{2.30}$$

for all possible values of the risk aversion coefficient, and where $b_k = 0$ means that all the elements in the K row vectors should be equal to zero.

Note that the main difference between (2.10) and (2.25) is the presence of the product of the information variables z_t and the returns on the initial set of K assets. Consequently, the disadvantage of this method is the dimensionality problem, since in case of L information variables and K initial assets, there are $(L+1)K+1$ regressors in the case that $N=1$. As in the unconditional case, all hypotheses in the conditional framework can be tested using standard Wald tests. However, this implies for (2.30) that $2+LK$ restrictions have to be tested.

Similar to the unconditional case, we can define the vector

$$\alpha_J^z(\eta) \equiv (\mu_r^z - \eta\iota_N) - B^z(\mu_R^z - \eta\iota_K) \tag{2.31}$$

as the conditional generalized Jensen measure. Under the assumption that B^z is a linear function of Z_t, the left hand side of (2.27) equals this conditional generalized Jensen measure. Moreover, a positive element in the vector of conditional generalized Jensen measures can be interpreted as outperformance of dynamic investment strategies by the mutual fund, while a negative value implies underperformance.

An alternative method to incorporate conditional information assumes that the expected returns are time-varying instead of B^z. In Chapter 3 of this thesis it is shown that under the assumption that μ_R^z is a linear function of the information variables, i.e. $\mu_R^z = \Gamma' Z_t$ where Γ' is of dimension $K \times (L+1)$, and that the conditional covariance matrices are constant over time, implying that B^z is time invariant, that a comparable regression framework as in (2.10), with $\Gamma' Z_t$ as regressors, can be used to test the hypothesis whether there is no extension of the investment set for only one value of the risk aversion coefficient or the hypothesis that this holds for all values of the risk aversion coefficient.

A third method to incorporate conditional information in performance evaluation involves the use of scaled returns $Z_t \otimes R_{t+1}$ (see, e.g. Bekeart and Urias [1996], Cochrane [1997]). These scaled returns can be interpreted as returns to dynamic investment strategies. For instance in the case of only one information variable, i.e. $L=1$, if a high value of z_t predicts that the returns are likely to be high in the next period, then the investor can invest in z_t units of R_{t+1} and the investor receives $z_t R_{t+1}$ the next period (see, e.g. Cochrane [1997]). The crucial difference with Ferson and Schadt [1996] is the introduction of LK additional assets in the investor's initial mean-variance optimization problem. It is straightforward to derive the optimal weight vector if it is taken into account that since the current price p_t of a return R_{t+1} is by definition equal to 1, the scaled returns $Z_t \otimes R_{t+1}$ will have an average price $q_K \equiv E[Z_t \otimes \iota_K]$.

Similar to the case of K initial assets, in case the investor cannot extend the efficient set by investing in the set of N mutual funds, it is straightforward to show that for a predetermined value of the zero beta rate η, the following expression holds for the last $N(L+1)$ elements of the extended set of assets:

$$\mu_r^z - \eta q_N = B^z(\mu_R^z - \eta q_K), \tag{2.32}$$

where μ_r^z and μ_R^z also include the scaled returns on the N mutual funds and the K initial assets respectively. The introduction of the scaled returns imply that the dimension of B^z in (2.32) changes to $(L+1)N \times (L+1)K$. If the expected returns in (2.32) are replaced by realized returns, a similar regression framework as before can be used to test for extension of the efficient for one particular risk aversion coefficient or for all possible risk aversion coefficients. However, the number of parameters to estimate and the number of restrictions to test is the main disadvantage of this method.

Recall that one of the possible reasons for the popularity of mutual funds is the potential special abilities of the fund managers. Thusfar we did not make the assumption that portfolio managers use superior information in their investment decision. However, in the mutual fund performance evaluation literature it is common to distinguish two kinds of special abilities which are based on the use of superior information (see, e.g. Admati, Bhattacharya, Pfleiderer and Ross [1986], Grinblatt and Titman [1989b]). First of all there is the ability to select the stocks with a higher expected return conditional on the managers' information than the publicly expected return, i.e. the so-called selection ability. For instance, the fund manager could have received an information signal that provides him the knowledge that the probability of a low return for an asset in the next period is much lower than expected by individual investors. Under the assumption that the manager interprets the signal in the right way, it can be expected that he will increase the weight of the asset in the mutual fund. A common assumption in the performance evaluation literature is that the behavior of the fund manager will not affect the publicly expected return of the asset, i.e. the behavior of the fund manager does not serve as a signal for individual investors.

The second special ability is to time the market. This means that a fund manager will increase (decrease) the sensitivity of the mutual fund for the market when the manager receives a signal that makes the expected return on the market index conditional on the managers' information higher (lower) than expected by individual investors. Note that this supposes time-varying expected returns. If present, both types of abilities make actively managed mutual funds an attractive investment product.

If expected returns and (co)variances are time-varying, it is straightforward to show that estimation of the misspecified model in (2.10) will be biased and can even lead to negative estimates of the unconditional generalized Jensen measure (see, e.g. Grinblatt and Titman [1989b]) even if $\alpha_j^z(\eta) = 0$ for all η. For instance, in the case that β_t is a linear function of z_t, and where high values of z_t predict that the returns are likely to be high in the next period, β in (2.10) will be overestimated and possibly cause a negative estimate for $\alpha_J(\eta)$. Alternatively, this can be interpreted as an omitted variable bias in equation (2.10).

In order to measure timing ability of fund managers, a number of methods are suggested in the literature. For instance, following the same analysis as Admati et al. [1986], and extending the first approach discussed above to incorporate conditional information, Ferson and Schadt [1996] assume that the elements in B^z are a linear function of the public information variables z_t and the private information signal $(R_{t+1}+\phi_t)$, where ϕ_t can be interpreted as a measurement error, i.e.

$$B_k^z(z_t,\phi_t) = c_{0k} + c'_{1k}z_t + c'_{2k}(R_{t+1}^k + \phi_t^k), \tag{2.33}$$

where $B_k^z(z_t,\phi_t)$ denotes the k-th element of the $1\times K$ dimensional row vector B^z, and $c'_{1k} \equiv (\tilde{c}_{1k}\ldots\tilde{c}_{lk})$ is a $1\times L$ row vector. Denote $c'_0 \equiv (c_{01}\ldots c_{0k})$, $c_1^z \equiv (c'_{11}z_t\ldots c'_{1k}z_t)$ and $c'_2 \equiv (c_{21}\ldots c_{2k})$. Substituting realized returns for conditional expected returns in (2.23) gives[1]

$$r_{t+1} = \alpha + c'_0R_{t+1} + c_1^zR_{t+1} + c'_2R_{t+1}^2 + v_{t+1}, \tag{2.34}$$

where $\alpha = \mu_r^z - B^z\mu_R^z$ and

$$v_{t+1} = (r_{t+1}-\mu_r^z) - B^z(R_{t+1}-\mu_R^z) + c'_2\phi_tR_{t+1}. \tag{2.35}$$

Equation (2.34) can be interpreted as the conditional version of the quadratic regression suggested by Treynor and Mazuy [1966]. Note that c_2 in (2.34) measures the effect of the private information signal, while $c_1^zR_{t+1}$ corrects for publicly available information.

Thusfar we presented a regression framework that can be used to answer the question whether investors can extend the efficient set by taking a position in a mutual fund. The framework presented in (2.33) to (2.35) is developed for measuring timing ability of mutual fund managers that is not based on publicly available information. Since squared return R_{t+1}^2 is not the return on an investment strategy, estimation of (2.34) does not directly answer the question whether there is an extension of investment set by taking a position in the fund under consideration relative to a prespecified set of benchmark assets. However, a positive estimate for c_2 in (2.34) can be interpreted as timing ability based on private information.

[1] Let R_{t+1}^2 denote squared returns and ϕ_tR_{t+1} reflect the element wise multiplication of ϕ_t and R_{t+1}.

2.2.5 Evaluating Performances of Conditional Strategies

We will now continue our discussion of empirical results, explicitly focussing on the conditional performance evaluation literature. As discussed in Section 2.2.4, simply estimating an unconditional performance evaluation model while conditional moments are time-varying may lead to biased estimates of the unconditional Jensen measure even if the benchmark assets are conditionally efficient. The unconditional performance evaluation model may suggest that the investor should take a short position in the mutual fund under consideration for an improvement in the risk-return trade-off. However, when we incorporate a set of economic variables that reflect the state of the economy, the mutual fund might show outperformance, indicating that the investor should actually take a long position in the fund under consideration.

In order to illustrate the size of this effect, Ferson and Schadt [1996] estimate the unconditional in (2.18) as well as the following conditional one-factor model

$$r_{i,t+1} = \alpha_i + b_0(r^m_{t+1}) + b'_1(z_t(r^m_{t+1})) + \varepsilon_{i,t+1}, \tag{2.36}$$

where $r_{i,t+1}$ is the return on mutual fund i in period $t+1$ in excess of the risk free rate $r_{f,t+1}$, r^m_{t+1} is the corresponding excess return on the market portfolio, and z_t is a vector of the following (predetermined) information variables that previous studies have shown to be useful for predicting stock returns: the dividend yield on the market portfolio, the yield on a short term Treasury Bill, the term spread (measured as the difference between the yield on a long and short term bond), a corporate bond yield spread (measured as the difference between the yield on low grade and high grade bonds), and a dummy for the month January. Recall that in the unconditional version of (2.36) the term $b'_1(z_t(r^m_{t+1}))$ is absent. In the model, a test for abnormal performance is equivalent to testing whether $\alpha_i = 0$. Moreover, a positive value for α_i indicates that the individual investor that is currently dynamically investing in the market portfolio and the risk free asset, can extend the mean-variance efficient investment set by taking a long position in the mutual fund under consideration, while a negative value indicates a short position in the fund. Alternatively, a test for $\alpha_i = 0$ can be interpreted as a test for the validity of the conditional CAPM as the right pricing model.

Over the sample period 1968-1990, Ferson and Schadt report for the unconditional version of (2.36) an average underperformance of 0.03% per month (i.e. 0.36% annually). Moreover, the mutual funds that are classified as 'maximum capital gains' even have an average underperformance of 0.96% annually. However, using the conditional one-factor model (2.36) the funds in the investment category 'maximum capital gains' appear to have an average outperformance of about 1.01%. This implies that dynamically investing investors whose initial portfolio is reflected by the market portfolio and the risk free asset can extend the efficient set by taking a long position in these kinds of mutual funds, while the unconditional model suggested that the investors should take a short position. Moreover, in case of the four-factor equivalent of (2.36), Ferson and Schadt find a comparable difference between the unconditional and conditional Jensen measure.

Since Ferson and Schadt [1996] do not test for the absence of abnormal performance under specific economic circumstances $\tilde{z}$, and moreover, they assume that a risk free asset is available, the hypothesis they test is in fact equivalent to (2.29), however they impose that $1 = b_0' \iota_K$ what implies that a part of the hypothesis already holds. Consequently, the zero beta rate η becomes irrelevant. The hypothesis that the product of the information variables with the market portfolio does not matter, is rejected for most of the funds in the sample. However, it appears that the corporate bond yield spread and the dummy for the month January have insignificant parameter estimates, indicating that these information variables have little prediction power. Alternatively, it can be said that there is no January effect (or anomaly) present in this data set.

As discussed in Section 2.2.4, in order to measure timing ability of fund managers, a quadratic term can be added to (2.36), which gives

$$r_{i,t+1} = \alpha_i + c_0(r^m_{t+1}) + c_1'(z_t(r^m_{t+1})) + c_2(r^m_{t+1})^2 + v_{i,t+1}, \tag{2.37}$$

where the returns are in excess of the risk free rate $r_{f,t+1}$. For a sample of internationally investing U.S. based mutual funds, Cumby and Glenn [1990] test for timing ability in an unconditional setting, i.e. the term $c_1'(z_t(r^m_{t+1}))$ is not present. Using the MSCI World index as market portfolio, they only find negative estimates for the timing coefficient c_2 over the sample period 1982 through 1988, indicating that mutual fund managers have a kind of perverse timing ability. This conclusion is confirmed by Ferson and Schadt [1996] for their sample of mutual funds and sample period. However, using the conditional framework (2.37), Ferson and Schadt mostly find positive estimates for c_2, indicating that mutual fund managers indeed have timing ability that is based on extra information.

2.2.6 Return-based Style Analysis

As discussed in Section 2.2.1, relative performance evaluation, i.e. the comparison of the return of a fund with the return on some appropriate benchmark, is the most popular performance measure in newspapers. In Section 2.2.1 we claimed that the actual investment style of a mutual fund does not necessarily correspond to the reported style (see, Brown and Goetzmann [1997]). Moreover, most mutual funds hold some cash as well. Both reasons make relative performance evaluation somewhat unfair, since mutual fund returns are compared with each other or with a benchmark asset that is supposed to reflect the mutual fund's investment style. Return-based style analysis, introduced by Sharpe [1992], can be used for obtaining an estimate of the fund's investment style and the exposure to certain asset classes. Examples of asset classes are for instance, growth stocks, value stocks, sector indices or market indices.

In order to estimate the investment style, the following regression equation can be estimated

$$r_{t+1} = a + \sum_{k=1}^{K} b_k R_{kt+1} + u_{t+1} \tag{2.38}$$

$$s.t. \sum_{k=1}^{K} b_k = 1 \tag{2.39a}$$

$$b_k \geq 0 \tag{2.39b}$$

where r_{t+1} denotes the return on a mutual fund, K is the number of asset classes, b_k reflects the sensitivity of the fund return for the asset class return R_{kt+1} and u_{t+1} is the idiosyncratic fund return, independent of all asset class returns. The fact that the exposures should sum to one implies that $\sum_k^K b_k R_{kt+1}$ can be interpreted as the return on a passive portfolio with the same style as the fund, while the positivity constraint reflects the short selling restrictions that are often present for mutual fund managers. Furthermore, the constant a is the average tracking error between the fund and the passive portfolio, and can therefore be interpreted as return due to selection.

Note that the style regression only works well when the fund return is highly correlated with the K asset classes, otherwise constraint (2.39a) might lead to misclassified styles (see, e.g. Fung and Hsieh [1997]). In contrast to Sharpe's style analysis, Brown and Goetzmann [1997] assume that a mutual fund is exposed to one asset class only. In Chapter 4 of this thesis we analyze the performance of a sample of Dutch mutual funds using style analysis. Moreover, we discuss the connection between style analysis and performance evaluation in more detail, and we present sufficient conditions under which relative performance evaluation is the appropriate method in detecting a fund manager's ability.

2.2.7 Performance Measurement Using Portfolio Holdings

Most of the papers discussed in this survey have in common that performance measurement is based on return data only. Since the end of the eighties, new datasets on mutual fund returns together with the corresponding fund holdings become slowly available. The frequency of the availability of the fund holdings determines the value of this extra information, and unfortunately at present often only say quarterly information on the holdings is available. Nevertheless it is of interest to consider the evaluation methods that have been developed to take the portfolio holdings into account (see, e.g. Grinblatt and Titman [1993], Daniel, Grinblatt, Titman and Wermers [1997]).

As claimed by Daniel et al. [1997], directly evaluating fund holdings has a number of advantages. The portfolio holdings of the funds can be applied to detect the investment style of the fund, such that the corresponding benchmark assets can be used for relative performance evaluation. This makes return-based style analysis as discussed in Section 2.2.6 not really necessary. Furthermore, the hypothetical returns that can be generated from the holdings are not affected by operating expenses of the fund. In that sense, performance evaluation using portfolio holdings is more suitable for detecting potential selection and timing ability of the fund managers. Due to the fact that we and many others do not have a dataset available that also contains the fund holdings at a regularly basis, we will in this survey and thesis not pay attention on performance measurement methods that incorporate fund holdings.

2.3 Persistence in Fund Performances

2.3.1 Introduction

Mutual funds often prominently advertise their past performance. When the past performance is good relative to other mutual funds or some benchmark asset, individual investors are influenced by these advertisements, investing more in funds with a good past performance (see, e.g. Sirri and Tufano [1997]). As claimed by Gruber [1996], if management ability exists, and is not included in the price of the mutual fund, then past performance should be predictive for future performance. Consequently, active mutual fund selection strategies can increase the expected return on a portfolio if mutual fund performance is really predictable.

The hypothesis that mutual funds with a good performance in this period will also have a good performance in the next period, implying a predictable pattern, is known as the hypothesis of persistence in performance. This terminology is used for time-varying expected returns as well as time-varying risk-adjusted expected returns. Moreover, fund managers that recently showed a good performance together with the presence of positive autocorrelation in mutual fund returns or risk-adjusted returns are referred to as having 'hot hands'.

Empirical studies often find that persistence in performance is a short-term phenomenon (see, e.g. Hendricks, Patel and Zeckhauser [1993], Wermers [1997]). A number of possible explanations for the disappearance of persistence in performance after some periods are:

1. the salaries of the fund managers and the fees that the funds charge rise, capitalizing recent successes.
2. once the reputation is established, mutual fund managers put less energy in selecting stocks
3. superior managers get bid away once a track record has been built
4. the flow of new money is too large, leading to fewer good investment opportunities per managed dollar

The remainder of this section is organized as follows. In Section 2.3.2 we will briefly discuss some methods used to measure persistence in performance, using much of the terminology of Section 2.2. Section 2.3.3 presents some empirical results of recent studies in performance persistence.

2.3.2 Measuring Persistence in Performance

The main question of interest in studies of performance persistence is whether past performance is to some extent informative about future performance. In the method most often applied, two periods are distinguished. In the first period, the so-called selection period, the mutual funds are ranked on the basis of returns or risk-adjusted returns. These rankings are split up into subgroups of mutual funds, and in the subsequent period, called the performance period, the performances of the subgroups are evaluated. Subsequently, it is examined whether the best performing mutual funds of the selection period continue to be the best in the performance period, indicating persistence in performance. More formally, let us define c_i as the rank order in the selection period of a performance measure x_i of mutual fund i divided by the total number of mutual funds available, N, i.e.

$$c_i = \frac{1}{N} \sum_{j=1}^{N} I(x_j < x_i) \;\; j \neq i, \tag{2.40}$$

where $I()$ is the so-called indicator function that equals one if the performance measure x_j of fund j, is smaller than the performance measure x_i of fund i. Note that c_i is between zero and one. Examples of performance measures of a mutual fund in the selection period are the generalized Jensen measure or the average return of a fund over the selection period.

In order to examine whether the rank c_i in the selection period is informative about a performance measure in the evaluation period, it can e.g. be tested whether

$$E[\alpha_{J,i}\,|\,c_i] > E[\alpha_{J,j}\,|\,c_j], \text{ with } c_i > c_j, \tag{2.41}$$

i.e. is the expected value of the generalized Jensen measure in the evaluation period larger given that the fund's rank order in the selection period is larger, or

$$E[r_{i,t+1}|\,c_i] > E[r_{j,t+1}|\,c_j], \text{ with } c_i > c_j, \tag{2.42}$$

i.e. is there a positive difference in the expected returns of fund i and fund j given that fund i was ranked higher in the selection period. If it is the case that the funds with the highest (lowest) rank in the selection period are also the best (worst) performing funds in the evaluation period, this can be interpreted as persistence in performance.

Note that testing of (2.41) is sensitive for the K factor mimicking portfolios included in the performance evaluation model. Recall that Carhart [1997a] uses a four-factor model for evaluating mutual fund returns, because Fama and French's [1993] three-factor model failed to explain cross-sectional variation in momentum sorted portfolios. Consequently, persistence in the Jensen measure $\alpha_{J,i}$ obtained from the Fama-French model can also imply that the fund successfully follows a momentum strategy, that leads to the observed persistence pattern. Moreover, recall from Section 2.2.2 that a positive value for the generalized Jensen measure indicates outperformance of the K benchmark assets by the mutual fund under consideration. Consequently, the question whether there is a predictable pattern in the generalized Jensen measure can be interpreted as looking for mutual funds that continue in outperforming the K benchmark assets over time. Therefore, similar to the empirical results in Section 2.2.3, the choice of the K benchmark assets determines the question that is answered in the persistence analysis.

Recall that the reason why conditional performance evaluation is advocated is that in recent studies on stock and bond returns it is shown that returns are predictable over time. Consequently, the assumption of constant expected returns and (co)variances that is made in unconditional performance evaluation is hard to justify. Christopherson, Ferson and Glassman [1998] examine the question whether conditional generalized Jensen measures (2.32) contain a persistent pattern, where the conditioning is upon a set of information variables supposed to reflect the state of the economy.

The procedure of examining whether a ranking is informative about future performance that has been sketched above has a non-parametric character. A parametric procedure for testing for persistence can be based on the following set-up. Consider a set of N mutual funds and assume that the conditional expected return of mutual fund i in period t can be written as

$$E[r_{it}|I_{t-1}] = \gamma_{i0} + \sum_{j=1}^{J} \gamma_{ij} r_{i,t-j} = \mu_i + \sum_{j=1}^{J} \gamma_{ij}(r_{i,t-j} - \mu_i), \tag{2.43}$$

where μ_i is the unconditional expected return. The coefficients γ_{ij} reflect persistence in return of fund i relative to its own unconditional mean, where positive coefficients and above average returns are usually interpreted as fund managers having 'hot hands'. Recall from Section 2.2 that a potential persistence pattern in (2.43) does not imply that individual investors can extend the efficient set by taking a position in the mutual fund under consideration. In order to answer that question, the covariance of the initial assets in portfolio with the set of mutual funds has to be taken into account. In (2.43) the focus is explicitly on the predictable pattern in mutual fund returns.

In order to test whether a predictable pattern in mutual fund returns is present, one can simply estimate the following regression equation

$$r_{it} = \gamma_{i0} + \sum_{j=1}^{J} \gamma_{ij} r_{i,t-j} + \varepsilon_{it}, \tag{2.44}$$

where $\varepsilon_{it} = (r_{it} - E[r_{it}|I_{t-1}])$ is the unexpected return of fund i in period t, and test whether the persistence coefficients γ_{ij} are significantly different from zero. Although (2.44) can simply be estimated for the N mutual funds separately, it is common practice to pool the returns of all funds and estimate a set of common persistence coefficients. This choice can be motivated by the fact that some mutual funds only have a very short history of fund returns available, implying inaccurate estimates, or because of the fact that the question of interest is whether fund returns are on average predictable.

In Chapter 5 of this thesis we discuss a number of approaches that have been suggested to estimate short-term persistence in mutual funds returns. Most of the approaches are based on estimating the following cross-sectional regressions

$$r_{it} = k_t + \sum_{j=1}^{J} \gamma_{jt} r_{i,t-j} + u_{it}, \tag{2.45}$$

where it is imposed that there is a time-varying homogeneous persistence pattern over the funds. Note that (2.45) also supposes a common time-varying mean, implying that the expected return on each of the funds is the same. This assumption might be very hard to justify in empirical applications. Moreover, it is straightforward to show that if there is variation in expected returns over the funds, estimating of (2.45) by Ordinary Least Squares will lead to a spurious

Table 2.1: **Summary Persistence in Performance** The results in the table are taken from Tables III and VI in Carhart [1997a]. Panel A reports average returns and estimated values for the generalized Jensen measures (column labelled 'Alpha') and exposure coefficients for the best (and subgroup top-third) and worst performing funds (and subgroup bottom-third) (deciles) formed on lagged one-year return. Panel B reports average returns and estimated values for the Jensen measure and exposure coefficients for the best and worst performing funds (deciles) formed on the Jensen measure in the selection period. In the table, VWRF is the excess return on the CRSP value-weighted market index, RMRF is the excess return on a value-weighted aggregate market proxy, SMB, HML and PR1YR are returns on value-weighted, zero-investment, factor mimicking portfolios for size, book-to-market and one-year momentum in stock returns.

Panel A : decile portfolios formed on lagged one-year return								
	Excess	one-factor model		four-factor model				
portfolio	Return	Alpha	VWRF	Alpha	RMRF	SMB	HML	PR1YR
(top)	0.75%	0.27%	1.08	-0.11%	0.91	0.72	-0.07	0.33
best	0.68%	0.22%	1.03	-0.12%	0.88	0.62	-0.05	0.29
worst	0.01%	-0.45%	1.02	-0.40%	0.93	0.32	-0.08	-0.09
(bottom)	-0.25%	-0.74%	1.05	-0.64%	0.98	0.32	-0.04	-0.17

Panel B : decile portfolios formed on four-factor Jensen measure							
	Excess	four-factor model					Exp
portfolio	Return	Alpha	RMRF	SMB	HML	PR1YR	Ratio
best	0.62%	0.02%	0.93	0.48	-0.14	0.14	1.13
worst	0.19%	-0.43%	0.93	0.49	-0.06	0.11	1.76

persistence pattern. A solution for this problem suggested in the literature is to subtract some estimate for the expected return on fund i, i.e. μ_i, from the left hand side variable in (2.45) (see, e.g. Jegadeesh [1990] and Hendricks, Patel and Zeckhauser [1993]). In Chapter 5 of this thesis we show that a number of estimators suggested for μ_i indeed solve the estimation problem, but generate another spurious pattern. Therefore, we suggest an alternative estimation method that is based on removing fixed individual effects in a dynamic panel data model.

2.3.3 Empirical Persistence Results

To illustrate the potential presence of persistence in performance, this section presents some empirical results of recent studies. Carhart [1997a] examines the persistence in performance of equity funds over the period 1962 through 1993. The mutual funds are sorted on the basis of one-year lagged returns or on the basis of the Jensen measure of the four-factor model given in (2.19). Moreover, these rankings are split up into deciles, implying that the mutual funds with rank order $c_i > 0.9$ form the portfolio of the best performing funds in the selection period, while the funds with rank order $c_i \leq 0.1$ form the portfolio of the worst performing funds. In the performance period, Carhart [1997a] evaluates the performances of the funds with the one-factor model given in (2.18) and the four-factor model given in (2.19). In Panel A of Table 2.1 we summarize his results for the best and worst performing deciles formed on lagged one-year returns. Moreover, we report the results for the top-third performing funds in the best decile and the results for the bottom-third performing funds of the worst decile.

It appears that best performing funds in the selection period outperform the worst performing funds with almost 0.70% (i.e. approximately 8% annually) in the performance period. The difference between the top performing and the bottom performing funds is even 1% per month. Note that the bottom performing funds have a negative excess return in the evaluation period, indicating that these funds continue to underperform the risk free asset. Moreover, the best performing funds have a positive one-factor Jensen measure, indicating that individual investors whose initial portfolio is reflected by the risk free asset and a value weighted market index can improve the risk-return trade-off by investing in these mutual funds. In case of the four-factor model, the worst performing funds have a negative Jensen measure. Alternatively, it can be said that the mutual funds underperform the four factor mimicking portfolios.

In order to explain the difference in performance between the best and the worst performing funds, or put differently, in order to explain the observed persistence pattern, Carhart looks at the exposure to the factors involved. In case of the one-factor model, it appears that the two deciles are almost equally exposed to the market portfolio (i.e. 1.03 vs 1.02), indicating that the one-factor model cannot explain the difference in returns. However, in case of the four-factor model it appears that the difference in exposure to size (SMB) and the momentum (PR1YR) portfolio explain a lot. The best performing mutual funds are much more sensitive for the performance of small stocks (i.e. 0.62 vs 0.32). Moreover, the best performing funds are positively correlated while the worst performing funds are negatively correlated with the momentum portfolio. Therefore, it can be concluded that exposure to common factors, such as a momentum factor, explain much of the observed persistence in mutual fund returns (see also Brown and Goetzmann [1995]). However, for the bottom performers there still remains some persistence, that cannot be fully explained by the differences in exposure to the common factors. A remarkable result is that the worst performing funds have a much higher expense ratio than the other nine deciles, i.e. 1.92% vs 1.22% annually (not reported in Table 2.1). Since there is no real difference in the size of the funds, there is not a clear reason for this difference (see also Elton, Gruber and Blake [1996]).

In Panel B of Table 2.1 we summarize Carhart's [1997a] results for the best and worst performing funds formed on the basis of the four-factor Jensen measure (estimated over a three-year period). The spread in the excess return is only 0.40% (i.e. approximately 5% annually), while the spread in the four-factor Jensen measure is 0.45% per month. The exposure to the four factor mimicking portfolios is almost equal for the best and worst performing funds, indicating that these factors do not explain the difference in performance. However, a part of the difference can be explained by the corresponding expense ratio of the two decile portfolios (i.e. 1.13% vs 1.76% annually). Looking at the funds in the decile with rank order in the range $0.1 \leq c_i \leq 0.2$, i.e. decile nine (not reported), it appears that there is a difference in the four factor Jensen measure of almost 0.25% with the worst performing funds (decile ten). Apparently, it is the case that the worst performing funds persist in performing bad and this cannot be explained by differences in exposure to the four factor mimicking portfolios.

In order to examine whether investors are aware of the observed pattern, i.e. returns and risk-adjusted returns are predictable to some extent, Gruber [1996] examines whether new cash flows to funds are predictable. It appears to be the case that the worst performing funds indeed have a negative new cash flow, while the best performing funds are rewarded with a positive new cash flow (see also Hendricks, Patel and Zeckhauser [1993], Sirri and Tufano [1997]). Moreover, it appears that investors can benefit from an active fund selection strategy. Gruber's result suggest that investors that withdraw the money invested in the worst performing funds and subsequently invest this money in new funds, earn a risk-adjusted return of almost 1% per year.

2.4 Survivorship Effects in Measuring Fund Performances and its Persistence

2.4.1 Introduction

Commonly used commercially available datasets often only contain the mutual funds that are still in existence. As argued by Brown, Goetzmann and Ross [1995] and by Hendricks, Patel and Zeckhauser [1997], empirical analyses based on currently traded funds that do not take this phenomenon into account can easily generate misleading inference due to this so-called survivorship bias. For instance, consider a mutual fund that takes a lot of risk. Such a fund will have a high probability of failure, in the sense that it realizes low returns. Low returns for a number of consecutive periods might the management of the fund let decide to merge the fund with another fund within the same investment company or to close down the fund completely and offering the investors the opportunity to switch to another fund. However, given that a mutual fund survived until the current period, it has probably realized high returns. Therefore, an analysis that does not address the problem of endogenous survival will easily generate an upward biased estimate of the expected return of a fund.

Survivorship bias is not only restricted to performance analysis of mutual funds, but is important in many more applications (see, e.g. Brown, Goetzmann and Ross [1995]). For instance, at the individual stock level, companies might go bankrupt, possibly leading to datasets that only contain successful companies. Event studies that measure the effect of a positive earnings announcement observe an upward drift in the cumulated return in excess of a risk-free rate. However, if only the funds that survived a financial distress position are taken into account in the analysis, some bias might be present. Closely related to this example is the equity premium puzzle. The fact that the historical premium provided by stocks over bonds is much larger than expected given a reasonable estimate of investors' risk aversion, might be due to the relative absence of returns on stocks that defaulted or merged because of poor performance.

The remainder of this section is structured as follows. In Section 2.4.2 we discuss the potential effect of survivorship on performance measures of mutual funds and its persistence. We discuss the claim that recent performance evaluation studies do not suffer from survivorship bias anymore, since all the funds that ceased to exist are taken into account until the moment of disappearance. Section 2.4.3 shows some empirical results of survivorship bias studies, and gives an impression of the size and impact of the bias in performance evaluation using traditional techniques.

2.4.2 Evaluating Survivorship Effects

The starting point in performance evaluation studies is the construction of a sample of mutual funds to be analyzed. In general one can distinguish two kinds of datasets that have been employed in performance evaluation studies. Early studies on performance evaluation often used a set that only contains the mutual funds that were still in existence at the moment of analysis, possibly leading to survivorship effects due to endogenous selection (see, e.g. Jensen [1968]). In the sequel, we will denote these samples as survivors only samples. Being aware of the potential survivorship effect, recent studies on performance evaluation employ samples that also contain the mutual funds that do not exist anymore (see, e.g. Brown and Goetzmann [1995], Carhart [1997a]). In this case the mutual funds are taken into account until the moment of disappearance. These samples are usually denoted as survivorship free samples. Most of the empirical performance evaluation studies discussed in the Sections 2.2.3, 2.2.5 and 2.3.3 employ such a survivorship free sample.

In Section 2.2.2 we discussed a number of performance measures that consider the question whether fund managers offer an investment product that was not attainable for individual investors before. Or put differently, whether the potential ability of the fund managers is a valid reason to invest in mutual funds. The simplest measure is the average return of a mutual fund or sample of funds. In order to examine what the effect of mutual fund disappearance is on performance evaluation and persistence in performance measures, knowledge of the survival or selection process is essential. Obviously, if survival of mutual funds is a random process, implying that the past record of fund returns, the investment style of the fund or other fund characteristics do not influence the probability that the fund is observed, performance evaluation will not be affected by survivorship.

Preluding on an example that will be extensively discussed in Chapter 6 of this thesis, we show that non-random survival generates survivorship effects in performance measures. Assume that there is a population of $i = 1, ..., M$ mutual funds. Moreover, two consecutive sample periods can be distinguished. A mutual fund realizes a good return with probability p_i, which can be interpreted as an indication of the ability of the fund manager. The cross-sectional mean of p_i is p. Which returns will be denoted as good is not important, but one could for instance distinguish between positive or negative returns. If the return of fund i in period t was good, then $r_{it} = \mu^H$, and μ^L otherwise. Consequently, the expected return on a mutual fund is

$$\mu_i = p_i\mu^H + (1 - p_i)\mu^L.$$

We assume that in period 1 all funds are observed, while in the second period a fund is observed with probability 1 if it had return $r_{it} = \mu^H$ in the first period, while it is only observed with probability q if it had return $r_{it} = \mu^L$. We index availability of fund i in the second period by $y_i = 1$.

The standard estimator for the expected return μ_i of fund i from a so-called 'survivorship free' sample is:

$$\hat{\mu}_i = \frac{r_{i1} + y_i r_{i2}}{1 + y_i}. \tag{2.46}$$

In Table 2.2 we report the possible outcomes for fund i in both periods together with the corresponding probabilities. By using the probabilities for each of the possible outcomes, it can be shown that

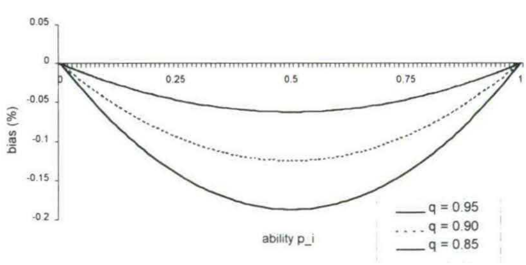
$$\begin{aligned} E[\hat{\mu}_i] - \mu_i &= (\frac{1}{2}p_i^2 - \frac{1}{2}p_i + (\frac{1}{2}p_i - \frac{1}{2}p_i^2)q)(\mu^H - \mu^L) \\ &= [\frac{1}{2}p_i(1 - q)(1 - p_i)](\mu^L - \mu^H), \end{aligned} \tag{2.47}$$

Table 2.2: **Non-random survival** The table reports the possible return realisation in the two consecutive sample periods with the corresponding probabilitity. The final two colums show the estimate for the expected return with the corresponding probability. NA indicates not available.

Period 1		Period 2		Estimate	
Prob.	Return	Prob.	Return	expected return	Prob.
p_1	r_{i1}	p_2	r_{i2}	$\hat{\mu}_i$	p
p_i	μ^H	p_i	μ^H	μ^H	p_i^2
p_i	μ^H	$(1-p_i)$	μ^L	$\frac{1}{2}(\mu^H+\mu^L)$	$p_i(1-p_i)$
$(1-p_i)$	μ^L	qp_i	μ^H	$\frac{1}{2}(\mu^H+\mu^L)$	$qp_i(1-p_i)$
$(1-p_i)$	μ^L	$q(1-p_i)$	μ^L	μ^L	$q(1-p_i)(1-p_i)$
$(1-p_i)$	μ^L	$(1-q)$	NA	μ^L	$(1-q)(1-p_i)$

what implies that a bias might be present in the estimator $\hat{\mu}_i$ for the expected return of fund i using a survivorship free sample. In Figure 2.1 we show the size of the bias (2.47) for three different values of survival probabilities q, where we assume that the difference between μ^H and μ^L is 10%.

Figure 2.1: Bias in expected return estimate. The figure shows the bias in the expected return estimate $\hat{\mu}_i$ for three different values of survival probabilities q, given that $\mu^H - \mu^L = 10\%$.



Obviously, in four rather special cases the bias will equal zero. First of all, the case where $\mu^H = \mu^L$, second, the case where $q = 1$, and finally the cases where $p_i = 0$ or $p_i = 1$. Moreover, it appears that the bias in the expected return estimate attains a maximum when the probability that a fund realizes a good return is 0.5, i.e. 0.12% in the case of survival probability $q = 0.9$. Consequently, even a survivorship free sample is not free of survivorship effects in the sense that the properties of standard estimators can be affected by the survival process.

In Chapter 6 of this thesis, we reconsider this example and also present an example concerning persistence in performance. Moreover, we show that in order to correct for survivorship bias, knowledge of the survival process is required. A probit model that specifies mutual fund survival probabilities can be estimated from a dataset that also contains the funds that ceased to exist, i.e. a survivorship free sample is a prerequisite for obtaining survivorship bias free results. The correction method that we propose involves the use of weights, which can simply be derived from the survival probabilities of the probit specification. The correction method will be illustrated by applying it to the example above.

2.4.3 The Size of Survivorship Effects

In order to illustrate the size and impact of survivorship effects in mutual fund performance evaluation using traditional techniques, we show some results of studies that take the problem of survivorship explicitly into account. Early studies on performance evaluation neglected fund attrition simply because the data on the disappeared funds was not available (see, e.g. Jensen [1968], Henriksson [1984]). Since the end of the eighties, mutual fund attrition receives more and more attention in the literature. One of the first studies that analyses the potential effect of survivorship on performance evaluation is Grinblatt and Titman [1989a]. A simple simulation analysis of the average return of a survivors sample and a survivorship free sample indicates that the effect varies between 0.1% and 0.3% per year. However, in the simulation experiment, it is unclear whether the size and impact of the survivorship effect varies for, for instance, mutual funds with different investment objectives.

Using all the equity funds that existed between 1971 through 1991, Malkiel [1995] provides some estimates of survivorship bias in average annual returns for funds with different objectives by comparing a survivors sample with a survivorship free sample. It appears that when the funds that ceased to exist are not taken into account that for funds with objective 'capital appreciation' the average return is 1.8% higher, while for funds with objective 'growth and income' the difference is only 0.4% per year. Note that both estimates of survivorship bias in average returns are larger than the 0.3% claimed by Grinblatt and Titman [1989a]. Another interesting result from Malkiel [1995] is that the nonsurviving funds have a significantly smaller return each year than the surviving mutual funds, indicating that low returns are a motive to let a fund merge or disappear.

Using a technique which they refer to as 'follow the money' Elton, Gruber and Blake [1995] provide estimates for survivorship bias in mutual fund performance evaluation using a one-factor model and a three-factor model. For the three-factor model Elton et al. report a difference between the average Jensen's alpha of the survivors sample and the survivorship free sample that varies between 0.70% and 0.90% per year dependent of the exact technique used, while for the one-factor model the difference varies between 0.30% and 0.70% per year. The main difference between the 'follow the money' technique and the inclusion of the funds in the sample until the moment of disappearance is that Elton et al. track all the funds that ceased to exist until the end of the sample period. If a fund merged with another fund, the hypothetical return of the mutual fund that ceased to exist is equal to that of the fund they merged with. Moreover, a policy change of a fund leads to a hypothetical return record that is the average of all the other funds.

Brown and Goetzmann [1995] focus explicitly on factors that affect mutual fund survival probabilities. A probit analysis of annual data shows that next to the record of past returns, the age of the fund, the size of the fund and the expense ratio are important factors in mutual fund survival probabilities. The older the fund the less likely it is that the fund will disappear the next year. Bigger funds have a lower probability of disappearance than small funds. The size and age effect are probably closely correlated with the effect of past returns, since high returns positively affect the investor's decision to invest in a fund (see, e.g. Sirri and Tufano [1997]). Bigger mutual funds probably had a good performance record, which attracts more money and which makes decision to close or merge a fund less likely. In Chapter 6 of this thesis, we extend the probit model of Brown and Goetzmann [1995] by allowing for aggregate macro-economic shocks that affect the survival of all mutual funds, such as for instance, bad returns on the stock market as a whole.

As mentioned in Section 2.4.2, if survival of mutual funds is non-random, but depends on endogenous factors such as past returns, this might seriously affect mutual fund performance evaluation. Most of the persistence in performance studies presented in Section 2.3 are conditional upon mutual fund survival. For instance, in order to compute the Jensen measure of a mutual fund, a fund has to be observed over a number of periods. This period, what Carhart [1997b] refers to as the look-ahead period, is linked to the performance evaluation method used.

Therefore, it might be the case that observed persistence patterns can be explained by a survivorship effect. In order to obtain insight in the size of the effect in persistence in performance studies, Brown, Goetzmann, Ibbotson and Ross [1992] performed a Monte Carlo simulation experiment. They generated random annual returns from the following one-factor model:

$$r_{i,t+1} - r_f = \beta_i(r^m_{t+1} - r_f) + \varepsilon_{i,t+1}, \tag{2.48}$$

where $r_{i,t+1}$ is the return on mutual fund i in period $t+1$, r^m_{t+1} is the corresponding return on the market portfolio and r_f is the return on a risk free asset. The equity premium and the index for systematic risk β_i are assumed to be normally distributed. The variance of the normally distributed idiosyncratic error term $\varepsilon_{i,t+1}$, is approximated by

$$\sigma_i^2 = k(1 - \beta_i)^2. \tag{2.49}$$

This relationship between non-systematic risk and the index for systematic risk is based on the observation that funds with a β close to unity have very low values of non-systematic risk, whereas funds that deviate from the market are less well diversified. The value of the constant of proportionality, k, is chosen such that the average value of R^2 is 0.90 across mutual funds, given the distribution of β and the assumed variance of the equity premium.

Using a simple survival process, where the bottom 5%, 10% or 20% of the mutual funds disappear from the sample each year, Brown et al. [1992] find a spurious persistence pattern in risk-adjusted returns. Funds with an above median performance in the selection period, have a more than 50% probability of being an above median performer in the evaluation period, where this probability is increasing in the cutoff percentage. In Chapter 6 of this thesis we reconsider this Monte Carlo simulation experiment, where we explicitly take a dynamic survival process into account that we obtained from a probit analysis to determine factors that affect mutual fund survival probabilities. In contrast to Brown et al., we split in the selection period the ranking of the funds in octiles. Our method closely corresponds to Hendricks, Patel and Zeckhauser [1997], who observe a spurious J shape persistence pattern using their survival process.

2.5 Contribution of this Thesis

In this survey we gave an extensive overview of mutual fund performance evaluation methods that are used to examine the reason why investors invest in mutual funds. In Chapter 3 of this thesis, we incorporate market frictions such as short selling constraints and transaction costs in a number of performance measures. We test for outperformance of the mutual funds in an unconditional as well as conditional framework. This chapter can therefore be seen as an extension of the framework discussed in Section 2.2 of this survey. In order to illustrate the effect of having a cash position on relative performance evaluation of mutual funds, we apply in Chapter 4 style analysis on a sample of Dutch mutual funds. This chapter extends style analysis in its application, but moreover, we show a direct connection between performance evaluation in a portfolio context and style analysis. The main contribution of Chapter 5 is that a number of approaches that have been suggested to estimate a potential persistence pattern in mutual fund returns, contain a bias. However, the sign of this bias does not generate a 'hot hands' phenomenon. Furthermore, we suggest an alternative estimation method that does not generate a spurious persistence pattern. The estimation methods are applied on a sample of U.S. based growth funds. Finally, we show in Chapter 6 that a survivorship free sample is not free of survivorship effects. A well-known approach in econometrics to handle endogenous samples is to model the survival process simultaneously with the phenomenon of interest. We propose to use these techniques to examine the size and the direction of the survivorship effect on performance evaluation and persistence in performance measures. The size of the bias that arises in persistence in performance studies is illustrated in a simulation experiment. Moreover, an extensive analysis of a sample of U.S. based mutual funds gives insight in the factors that affect mutual fund survival probabilities. Furthermore, we propose a weighting procedure based upon probit regressions that can be used to correct for the survivorship bias that might arise in performance evaluation using traditional techniques. The correction method is used to estimate persistence in performance of U.S.-based growth, aggressive growth and income funds.

Appendix 2.A

Optimal Portfolio Choice and Adjusting Weights

In this appendix we discuss the basic elements of Modern Portfolio Theory that are required to obtain and interpret the performance measures in the main text. A closely related discussion was presented before by DeRoon [1997]. Consider a mean-variance optimizing investor that chooses his portfolio from a set of K assets. Let the expectation and the covariance matrix of the K-dimensional return vector R_{t+1} be given by μ_R and Σ_{RR} respectively. The investor's utility function is of the form $f(\tilde{w}_R'\mu_R, \tilde{w}_R'\Sigma_{RR}\tilde{w}_R)$,where $\tilde{w}_R$ is the vector of portfolio weights and the function f is increasing in the first argument and decreasing in the second. The mean-variance problem yields the following Lagrangian:

$$max_{\{w\}}\ \mathrm{L} = f(\tilde{w}_R'\mu_R, \tilde{w}_R'\Sigma_{RR}\tilde{w}_R) - \eta(\tilde{w}_R'\iota_K - 1), \tag{A.1}$$

where ι_K is a K-vector of ones. Differentiating (A.1) gives the first order conditions:

$$f_1\mu_R + 2f_2\Sigma_{RR}\tilde{w}_R - \eta\iota_K \ = \ 0 \tag{A.2a}$$
$$\tilde{w}_R'\iota_K - 1 \ = \ 0, \tag{A.2b}$$

where f_1 and f_2 are the partial derivatives of the utility function with respect to its first and second argument. From the first order conditions it follows that the optimal weight vector w_R^* is determined by

$$\tilde{w}_R^* = \tilde{\gamma}^{-1}\Sigma_{RR}^{-1}(\mu_R - \eta\iota_K), \tag{A.3}$$

where $\tilde{\gamma}^{-1} \equiv -\frac{f_1}{2f_2}$ is the investor's risk aversion coefficient. From the first order conditions it also follows that $\tilde{\gamma} = \mu_R'\Sigma_{RR}^{-1}\iota_K - \iota_K'\Sigma_{RR}^{-1}\iota_K\,\eta$, implying that mean-variance efficient portfolios are uniquely determined when either the risk aversion coefficient $\tilde{\gamma}$ or Lagrange multiplier η is known. It is straightforward to show that for a given mean-variance efficient portfolio $\tilde{w}_R^*$, the Lagrange multiplier η equals the expected return on the zero-beta portfolio of $\tilde{w}_R^*$,which can be obtained as the intercept of the line tangent to the mean-variance frontier at $\tilde{w}_R^*$.

Suppose now that the investor takes a set of N mutual funds with N-dimensional return vector r_{t+1} into account. The expected return and (co) variances of these mutual funds are given by μ_r and Σ_{rr} respectively. The covariance with the set of initial assets is given by Σ_{rR}. Recall that we refer to the extended set of assets when a subscript is absent in the notation. It is straightforward to show that for a given value of η, the optimal weight vector w^* for the extended set of assets can be written as

$$w^* = \begin{pmatrix} w_R \\ w_r \end{pmatrix} = \gamma^{-1} \begin{pmatrix} \Sigma_{RR} & \Sigma_{Rr} \\ \Sigma_{rR} & \Sigma_{rr} \end{pmatrix}^{-1} \begin{pmatrix} \mu_R - \eta\iota_K \\ \mu_r - \eta\iota_N \end{pmatrix}, \tag{A.4}$$

which can easily be rewritten (using partitioned inverses) as

$$\begin{aligned} \begin{pmatrix} w_R \\ w_r \end{pmatrix} &= \gamma^{-1} \begin{pmatrix} \Sigma_{RR}^{-1} + B'\Sigma^{rr}B & -B'\Sigma^{rr} \\ -\Sigma^{rr}B & \Sigma^{rr} \end{pmatrix} \begin{pmatrix} \mu_R - \eta\iota_K \\ \mu_r - \eta\iota_N \end{pmatrix} \Leftrightarrow \\ w_r &= \gamma^{-1}\Sigma^{rr}(\mu_r - \eta\iota_N - B'(\mu_R - \eta\iota_K)) & \text{(A.5a)} \\ w_R &= \tilde{\gamma}\gamma^{-1}\tilde{w}_R^* - B'w_r, & \text{(A.5b)} \end{aligned}$$

where $B \equiv \Sigma_{rR}\Sigma_{RR}^{-1}$ and $\Sigma^{rr} \equiv (\Sigma_{rr} - \Sigma_{rR}\Sigma_{RR}^{-1}\Sigma_{Rr})^{-1}$. The part between parentheses on the right hand side of (A.5a) is also known as the generalized Jensen measure $\alpha_J(\eta)$, where η is the zero beta rate corresponding to the investor's initial portfolio of K assets. However, note that (A.5b) contains two different risk aversion coefficients. First of all $\tilde{\gamma}$, corresponding to the initial set of assets, and second, γ corresponding to the extended set of $N + K$ assets, which makes (A.5b) rather difficult to interpret.

The Sharpe ratio of a mean-variance efficient portfolio w_R^* is defined as

$$\theta_K(\eta) = \frac{w_R^{*\prime}\mu_R - \eta}{\sqrt{w_R^{*\prime}\Sigma_{RR}w_R^*}}, \tag{A.6}$$

i.e. the expected excess return divided by the standard deviation of the portfolio return. Using (A.3) this Sharpe ratio can easily be rewritten as

$$\theta_K(\eta) = ((\mu_R - \eta\iota_K)'\Sigma_{RR}^{-1}(\mu_R - \eta\iota_K))^{0.5}. \tag{A.7}$$

Similar expressions can be derived for the Sharpe ratio of mean-variance efficient portfolios of the extended set of assets, referred to as $\theta_{N+K}(\eta)$. Now define $\tilde{m} \equiv \tilde{w}_R^{*\prime}\mu_R$ and $m \equiv w^{*\prime}\mu$ as the expected returns on the investor's optimal portfolio for the initial and extended set of assets respectively. Then it is straightforward to show that by substituting (A.3) into the denominator of (A.6), the following relationship between the squared Sharpe ratio and the risk aversion coefficient holds:

$$\tilde{\gamma} = \frac{\theta_K^2(\eta)}{\tilde{m} - \eta}. \tag{A.8}$$

A similar relationship can also be derived for the squared Sharpe ratio and the risk aversion coefficient γ of the extended investment problem. Substituting the expressions for $\tilde{\gamma}$ and γ into (A.5a) and (A.5b) gives the new optimal weights of the extended investment problem without the two different risk aversion coefficients:

$$w_r = \left(\frac{m-\eta}{\theta^2_{N+K}(\eta)}\right)\Sigma^{rr}(\alpha_J(\eta)) \tag{A.9}$$

$$w_R = \left(\frac{\theta^2_K(\eta)}{\theta^2_{N+K}(\eta)}\right)\left(\frac{m-\eta}{\tilde{m}-\eta}\right)\tilde{w}_R - B'w_r. \tag{A.10}$$

Furthermore, by writing

$$\theta^2_{N+K}(\eta) = \gamma(m-\eta) = (\mu - \eta\iota)'\Sigma^{-1}(\mu - \eta\iota), \tag{A.11}$$

and using the partioned inverse, it can be shown that

$$\theta^2_{N+K}(\eta) = \theta^2_K(\eta) + \alpha_J(\eta)'\Sigma^{rr}\alpha_J(\eta). \tag{A.12}$$

Note that in analogy to (A.1), the first order conditions of the extended mean-variance problem (A.4) imply that

$$\gamma = \mu'\Sigma^{-1}\iota - \iota'\Sigma^{-1}\iota\,\eta, \tag{A.13}$$

which can easily be rewritten as (using the partioned inverse)

$$\gamma = \tilde{\gamma} + \alpha_J(\eta)'\Sigma^{rr}(\iota_N - B\iota_K). \tag{A.14}$$

By substituting (A.14) and using the expressions for γ and $\tilde{\gamma}$ given in (A.8), it is straightforward to show that (A.9) and (A.10) can be rewritten as

$$w_r = \left(\frac{\tilde{m}-\eta}{\theta^2_K(\eta) + (\tilde{m}-\eta)\alpha_J(\eta)'\Sigma^{rr}(\iota_N - B\iota_K)}\right)\Sigma^{rr}(\alpha_J(\eta)) \tag{A.15}$$

and

$$w_R = \left(\frac{\theta^2_K(\eta)}{\theta^2_K(\eta) + (\tilde{m}-\eta)\alpha_J(\eta)'\Sigma^{rr}(\iota_N - B\iota_K)}\right)\tilde{w}_R - B'w_r. \tag{A.16}$$

Chapter 3

Performance Analysis of International Mutual Funds Incorporating Market Frictions

In this chapter we analyze the performance of internationally investing U.S.-based mutual funds, correcting for market frictions such as short sell constraints and transaction costs using a variety of performance measures. We first of all show that for a number of funds Jensen's α is significantly positive if market frictions are ignored. Subsequently we show that the evidence of outperformance is robust to measuring performance with respect to an international asset pricing model with three country portfolios and a currency portfolio as the factor mimicking portfolios. As is well known by now these performance measures can alternatively be interpreted as tests of the hypothesis that the supposed factor mimicking portfolios span the efficient frontier of these portfolios and the mutual fund, i.e. of the hypothesis that mean-variance agents can not improve their risk return trade-off by also investing in the mutual fund.

3.1 Introduction

The empirical literature on performance evaluation of mutual funds concentrates on the question whether fund managers have special abilities in composing a portfolio (e.g., Jensen [1968]), which could provide investors with superior returns. The issue of performance measurement is closely related to the question whether investors can improve their portfolio's risk-return trade-off when additional assets are taken into account (see, e.g. Jobson and Korkie [1989], Chen and Knez [1996]). Performance measurement requires a pricing model in order to define outperformance of efficient benchmark portfolios. On the contrary, a test for a shift in the mean-variance frontier by including additional assets can do without a pricing model since it only starts with the assumption that the investor already holds an efficient portfolio of the benchmark assets. This implies that performance evaluation of mutual funds is equivalent to a test for diversification benefits under the assumption that investors hold efficient combinations of benchmark assets that correspond to the pricing model used. If no mean-variance optimizing investor can significantly extend the investment set by considering a set of additional assets then there is mean-variance spanning, as defined by Huberman and Kandel [1987]. If there is only a particular group of investors that cannot extend the efficient set by adding an additional asset then there is intersection. In the latter case the mean-variance frontier of the original assets in portfolio and the mean-variance frontier of the extended portfolio have one point in common.

An important shortcoming of many tests for mean-variance spanning proposed in the literature is the supposed absence of market frictions. When buying assets, investors are confronted with transaction costs. In particular, investors considering international diversification have to deal with high transaction costs. For these investors internationally diversified mutual funds are an alternative for obtaining a highly diversified portfolio (see Cumby and Glenn [1990]). However, mutual funds have operating expenses such as management fees, administrative costs, advisory fees and marketing costs which are deducted from the fund's assets. Although in performance evaluation studies the difference between returns before expenses and returns after expenses is taken into account (see, e.g. Malkiel [1995]), the load fees charged by some of the mutual funds are usually ignored.

Only recently, tests for mean-variance spanning have been extended to take frictions, such as transaction costs and short selling restrictions, into account (see Hansen, Heaton and Luttmer [1995], Luttmer [1996] and DeRoon, Nijman and Werker [1998]). In this chapter we employ a sample of internationally investing mutual funds, and we will show that for a large number of the funds the hypothesis of mean-variance spanning will not be rejected if short selling restrictions are incorporated, whereas in the case without market frictions, mean-variance spanning is rejected for most of these funds. Furthermore, we will show that incorporating load fees for mutual funds substantially affects the diversification benefits that investors can realize by including internationally investing mutual funds in their portfolio. This chapter extends the paper from DeRoon, Nijman and Werker [1998] in its application to mutual fund performance evaluation. Moreover, we incorporate market frictions in conditional performance evaluation, and we show how to test two-sided inequality restrictions that arise in mean-variance spanning tests where transaction costs are incorporated.

The remainder of this chapter is organized as follows. In Section 3.2 we show the relationship between performance evaluation and testing for diversification benefits by adding an asset to the initial portfolio. Furthermore, we present our sample of internationally investing mutual funds, and discuss some previous empirical results on performance measurement of mutual funds. In Section 3.3 we show how market frictions such as short sales restrictions and transaction costs can be incorporated in performance evaluation. Section 3.4 presents the tests for mean-variance spanning in case of a frictionless market as well as when market frictions are incorporated. The empirical results show that the possible diversification benefits by including internationally investing mutual funds are seriously affected by short selling constraints on some of the assets under consideration. In conditional performance measurement studies some predetermined information variables, that can be used to predict stock returns, are explicitly taken into account in evaluating mutual fund performances. In Section 3.5 we present the tests as well as the empirical results for potential diversification benefits of mutual funds in a frictionless as well as a market with frictions when we allow for time-varying expected returns. Finally, Section 3.6 concludes.

3.2 Performance Analysis of US based International Mutual Funds in a Frictionless Market

For a risk-averse optimizing agent, the decision to invest in an internationally diversified mutual fund depends on the question whether a fund manager is able to extend the investor's efficient set of assets. The superior risk-return trade-off that a mutual fund potentially provides due to timing or selection ability of the fund manager, will be the motive to add a fund to the initial portfolio of assets. Suppose a mean-variance investor considers to extend his initial efficient set of K assets by adding a set of N internationally investing mutual funds. The gross return vector r_{t+1}, represents the returns of the mutual funds after operating expenses. The gross returns for the K benchmark assets are denoted by the vector R_{t+1}. In a frictionless market, where the Law of One Price holds, there exists a stochastic discount factor M_{t+1} such that

$$E[M_{t+1}R_{t+1}|\, I_t] = \iota_K, \tag{3.1}$$

where ι_K is a K-vector of ones and I_t is the public information set available at time t. In the Sections 3.2, 3.3 and 3.4, we assume that the expected returns on the assets and the corresponding (co)variances are constant over time. Extensions to the conditional version of (3.1) will be implemented in Section 3.5, following Ferson and Schadt [1996].

Since we consider a mean-variance optimizing investor, a stochastic discount factor M_{t+1} is a linear function of the K asset returns. Moreover, as shown by Hansen and Jagannathan [1991], the mean-variance stochastic discount factor $m\,(v)_{t+1}$ given by

$$m\,(v)_{t+1} = v + \alpha(v)'(R_{t+1} - E[R_{t+1}]), \tag{3.2}$$

with

$$\alpha(v) = Var[R_{t+1}]^{-1}(\iota_K - vE[R_{t+1}]),$$

has the lowest variance of all stochastic discount factors with expectation v, that price R_{t+1} correctly. It is straightforward to show that the zero beta rate[1] corresponding to the mean-variance investor's optimal portfolio is equal to $1/v$, i.e. the inverse of the expectation of the stochastic discount factor. A mean-variance optimizing investor cannot extend the efficient set by investing in the mutual fund under consideration if the stochastic discount factor given in (3.2) also prices the mutual fund's return r_{t+1} correctly (see, e.g. Bekeart and Urias [1996], DeRoon, Nijman and Werker [1996]). This can be seen easily if we recognize that (3.1) can be interpreted as the first order conditions of an investors portfolio problem.

As mentioned, the question whether there is a shift in the efficient frontier by extending the investment set is closely related to performance measurement (see, e.g. Chen and Knez [1996]). A well known measure of the performance of a mutual fund is the generalized Jensen [1968] measure. It can be obtained as the intercept in a regression of the excess return[2] of the mutual fund on the excess returns of some benchmark portfolios and a constant. However, in order to evaluate the performance of a mutual fund a pricing model is required that specifies the set of K efficient benchmark portfolios that span the mean-variance frontier. For instance, under the assumption that the CAPM is the pricing model, the so-called market portfolio with return R_{t+1}^m and the risk free deposit are the efficient benchmark portfolios.

Since a minimum variance stochastic discount factor is linear in the returns of the benchmark assets, it is straightforward to show that the performance of a fund relative to the set of benchmark assets can be measured by

$$\lambda(v) = E[m(v)_{t+1}r_{t+1}] - 1 = v\alpha_J(v), \tag{3.3}$$

where $\alpha_J(v)$ is the generalized Jensen measure. A similar relationship can be derived for a multifactor pricing model, implying a stochastic discount factor $m(v)_{t+1}$ that is linear in the factor mimicking portfolios (see, e.g. Fama [1996]). The performance measure (3.3) indicates that an investor can improve the risk-return trade-off by buying the mutual fund if $\lambda(v) > 0$, and selling the fund, i.e. taking a short position, if $\lambda(v) < 0$. The case where $\lambda(v) = 0$ corresponds to no diversification benefits by including this fund into the portfolio. If $\lambda(v) = 0$ for exactly one value of v, this is equivalent with intersection of the extended and initial mean-variance frontiers at the point where the investor's optimal portfolio is located. Furthermore,

[1] The zero beta rate of a portfolio can be obtained as the intercept of the line tangent to the mean-variance frontier in the point where the investor's optimal portfolio is located.

[2] An excess return is the return in excess of a risk free rate (if available) or some zero beta rate.

if $\lambda(v) = 0$ holds for all possible v then the extended investment set is spanned by the original K benchmark assets, corresponding to the case where the extended and initial mean-variance frontier coincide.

In this chapter, we examine whether internationally investing U.S. based mutual funds can extend the efficient investment set of a U.S. investor. We employ a sample of eighteen internationally investing open-end mutual funds over the period 1982 through 1994. The mutual fund data are obtained from the Morningstar Mutual Fund Database. Morningstar reports information about all open-end mutual funds on a monthly basis. The mutual funds in our sample have as investment objective 'foreign' or 'world', and exist over the whole sample period. Our sample is comparable with the sample of Cumby and Glenn [1990], studying the performance of a sample of fifteen U.S. based internationally diversified mutual funds over the period January 1982 through June 1988.

Since performance evaluation depends on the choice of the set of benchmark portfolios, we consider in this chapter two sets of benchmark assets. The first set of benchmark assets is equivalent to the one employed by Cumby and Glenn [1990] and consists of the Morgan Stanley World index and an equally weighted portfolio of Eurocurrency Deposits[3] to reflect a currency hedge portfolio. This set of benchmark assets can be interpreted as the initial portfolio of a group of investors that have a widely diversified international portfolio with predetermined country weight allocation corresponding to the market capitalization of the individual countries and who consider to extend their portfolio with an internationally investing mutual fund. The second set of benchmark assets represents the initial portfolio of investors that currently invest efficiently in the US, European and Japanese stock indices as well as the Eurocurrency Deposits. The benchmark assets used are the Morgan Stanley Capital Market Indices (MSCI) for the USA, Europe and Japan, obtained from Datastream. Based on the claim that in international asset pricing the asset returns are better described by multifactor models than by single index models (see, e.g. Korajczyk and Viallet [1989]), the two sets of benchmark assets can alternatively be interpreted as tests whether a two or four factor model prices the mutual funds under consideration.

In Panel A of Table 3.1 we present some summary statistics for the sample of eighteen mutual funds that we employ. Note the variation in the front loads that the funds charge: five mutual funds can be marked as no-load funds while nine internationally investing funds charge more than 5.75% for a position in the mutual fund. The average returns for the no-load funds do not appear to be different from the funds that charge a load fee. Furthermore, it seems that

[3] The currencies in this portfolio are the Canadian dollar, the Deutsche mark, The Dutch guilder, the French franc, the Japenese yen, the pound Sterling and the Swiss franc.

Table 3.1: **Summary Statistics.** Panel A of this table contains some summary statistics for a sample of eighteen mutual funds. The average monthly return (corrected for operating expenses) and standard deviation, are calculated over the period 1982 through 1994. The column 'Net Assets' is the size of the fund in million dollars as reported at the end of 1994. The column 'Expense Ratio' reports the average monthly percentage that a fund took out of its assets for operating expenses over the period 1982 through 1994. The column 'Front Load' shows a one-time deduction the funds charge for an investment made into the fund. The column labelled 'Correl World' shows the correlation between the mutual fund and the MSCI World index. Panel B of the table shows summary statistics for the benchmark indices.

Panel A							
Mutual Fund	Average Return (%)	Stand. Dev. (%)	Net Assets (mln $)	Front Load (%)	Expense Ratio (%)	Correl. World	
Alliance Global Sm.	0.84	6.18	53.80	4.25	0.14	0.69	
Alliance Intl.	1.24	5.31	167.40	4.25	0.13	0.82	
Bailard,Biehl Intl.	0.99	5.21	116.20	0.00	0.09	0.84	
First Invest Global	1.14	5.23	206.80	6.25	0.15	0.78	
Kemper Intl.	1.14	4.60	323.40	5.75	0.11	0.82	
Lexington Wrd. Wide	0.94	5.26	261.80	0.00	0.13	0.69	
New Perspective	1.29	3.94	6560.50	5.75	0.06	0.87	
Oppenheimer Global	1.29	5.28	1809.50	5.75	0.12	0.81	
Phoenix World Opp.	0.82	6.09	122.50	4.75	0.12	0.67	
Pilot Kleinwort Intl.	1.08	5.07	30.00	4.50	0.15	0.85	
Putnam Global Gr.	1.31	4.27	1427.70	5.75	0.11	0.89	
Scudder Intl.	1.25	4.63	2131.80	0.00	0.10	0.86	
T. Rowe Price Intl.	1.35	4.76	5465.60	0.00	0.09	0.87	
Templeton Gr.	1.25	4.10	5727.70	5.75	0.13	0.79	
Templeton Sm. Cmp.	1.27	4.47	1253.00	5.75	0.08	0.73	
Templeton Wrd.	1.24	4.11	5123.40	5.75	0.06	0.79	
United Intl. Gr.	1.27	4.26	603.00	5.75	0.09	0.86	
Vanguard Intl. Gr.	1.41	4.99	2755.80	0.00	0.06	0.85	
Panel B							
Benchmark Indices	Average Return	Stand. Dev.	Correlations World	Eurocur	USA	Europe	Japan
World	1.26	4.28	1.00	0.28	0.76	0.82	0.77
Eurocur	0.84	2.82		1.00	-0.07	0.40	0.40
USA	1.25	4.34			1.00	0.59	0.26
Europe	1.41	4.88				1.00	0.51
Japan	1.53	7.57					1.00

the expense ratio of a fund is negatively correlated with the size of the fund. This can probably be explained by the fixed costs involved in managing a mutual fund. Moreover, since we defined returns as returns after operating expenses, a high expense ratio influences the average return of the fund. Panel B of Table 3.1 contains information about the benchmark portfolios. It appears that the Morgan Stanley Japan index realized the highest average return but also involves the highest risk as measured by the standard deviation. Furthermore, the returns on the USA, Europe and Japan stock indices are highly correlated, as expected, with the return on the Morgan Stanley World index.

For both sets of benchmark assets defined above, Table 3.2 contains the outcomes for the performance measure (3.3) for the case where the zero beta rate $\frac{1}{v}$ is set equal to the average monthly return on the one-month Tbill over the period 1982-1994, i.e. 0.53%. It appears that in

Table 3.2: **Generalized Jensen measure.** The table reports the generalized Jensen measure for two sets of benchmark assets. The first set consists of the Morgan Stanley World index and an equally weighted portfolio of Eurocurrency Deposits, and the second set contains the Morgan Stanley USA, Europe and Japan indices and an equally weighted portfolio of Eurocurrency Deposists. The zero beta rate $1/v$ is set equal to average monthly return on the one-month Tbill over the period 1982-1994, i.e. 0.53. The columns 'Returns after expenses' show the Jensen measure calculated with returns corrected for operating expenses, while the columns 'Returns before expenses' reports the Jensen measure before the fund's expenses are subtracted from its net assets. The results are based on monthly observations for the sample January 1982 through December 1994. Standard errors are reported in parentheses.

Mutual Fund	Generalized Jensen Measure			
	Two Bench. Assets		Four Bench. Assets	
	Returns after expenses	before expenses	Returns after expenses	Returns before expenses
Alliance Global Sm.	-0.29 (0.33)	-0.16 (0.33)	-0.57 (0.24)	-0.44 (0.24)
Alliance Intl.	-0.05 (0.25)	0.08 (0.25)	-0.17 (0.21)	-0.04 (0.21)
Bailard,Biehl Intl.	-0.32 (0.23)	-0.22 (0.23)	-0.36 (0.19)	-0.27 (0.19)
First Invest Global	-0.07 (0.27)	0.08 (0.27)	-0.18 (0.26)	-0.04 (0.26)
Kemper Intl.	-0.05 (0.22)	0.06 (0.22)	-0.14 (0.17)	-0.03 (0.17)
Lexington Wrd. Wide	-0.11 (0.28)	0.02 (0.28)	-0.33 (0.22)	-0.21 (0.22)
New Perspective	0.20 (0.16)	0.27 (0.16)	0.04 (0.12)	0.10 (0.12)
Oppenheimer Global	0.04 (0.26)	0.16 (0.26)	-0.12 (0.23)	-0.00 (0.23)
Phoenix Wrd. Opp.	-0.31 (0.35)	-0.19 (0.35)	-0.57 (0.28)	-0.45 (0.28)
Pilot Kleinwort Intl.	-0.20 (0.22)	-0.05 (0.22)	-0.30 (0.18)	-0.15 (0.18)
Putnam Global Gr.	0.13 (0.16)	0.24 (0.16)	-0.02 (0.12)	0.09 (0.12)
Scudder Intl.	0.02 (0.19)	0.12 (0.19)	-0.09 (0.15)	0.01 (0.15)
T. Rowe Price Intl.	0.08 (0.19)	0.17 (0.19)	-0.02 (0.14)	0.07 (0.14)
Templeton Gr.	0.23 (0.19)	0.36 (0.19)	0.06 (0.14)	0.18 (0.14)
Templeton Sm. Cmp.	0.27 (0.22)	0.35 (0.22)	0.09 (0.18)	0.17 (0.18)
Templeton Wrd.	0.22 (0.18)	0.28 (0.18)	0.03 (0.13)	0.09 (0.13)
United Intl. Gr.	0.11 (0.18)	0.20 (0.18)	0.01 (0.17)	0.10 (0.17)
Vanguard Intl. Gr.	0.12 (0.21)	0.18 (0.21)	0.04 (0.17)	0.10 (0.17)
average	0.00	0.11	-0.14	-0.04

case of returns after operating expenses, ten mutual funds have a positive performance measure in the two benchmark case, while in the four benchmark case only six funds have a positive value in our sample, none of them significant at the 5% level. The outperformance with respect to the two benchmark case of the best performing funds is in the order of magnitude of 0.25% per month, i.e. 3.00% annually. Our outcomes are in accordance with the results of Cumby and Glenn [1990], who find four out of fifteen internationally investing mutual funds with positive Jensen measures. Recall that the Jensen measure corresponds with a stochastic discount factor M_{t+1} that is linear in the benchmark assets. Consequently, the outcome of the Jensen measure can be interpreted in light of (3.3). This means that investors who already hold an efficient portfolio in the case of four benchmark assets, can only improve the risk-return trade-off by taking a short position in most of the mutual funds under consideration. However, it is important to note that taking a short position in mutual funds is almost impossible for most investors.

As noted by for instance Malkiel [1995], underperformance with respect to a set of benchmark portfolios does not mean that fund managers do not have special abilities in stock selection. Since mutual funds have operating expenses, reported as the expense ratio of the fund, that are deducted from the fund's assets, the performance evaluation of returns before expenses may give some indication for special abilities such as timing or selection ability. However, it has to be noted that the general investing public cannot benefit from these superior returns if the costs for obtaining this extra information are too high. Malkiel evaluates the returns from equity mutual funds over the period 1971 through 1991 by assuming the CAPM as the pricing model. In case of returns before expenses, he finds that mutual fund managers outperform the market portfolio with +0.18% on a monthly basis, i.e. about 2.00% annually, whereas underperformance dominates after expenses. In case of returns before expenses, we find in our sample of mutual funds that the performance measures for nine mutual funds are positive with respect to the four benchmark case (average: -0.04% monthly), while in the two benchmark case even fourteen out of eighteen funds have a positive performance measure (average: 0.11% monthly). Results similar to those of Malkiel are obtained by Carhart [1997a], and by Daniel, Grinblatt, Titman and Wermers [1997]. Note that all these papers ignore market frictions. The techniques that we employ to measure the performance of internationally investing mutual funds can be extended to the case of domestic equity portfolios e.g. by imposing short sell restrictions on small as well as large, and high book to market as well as low book to market stocks.

3.3 Performance Analysis in case of Market Frictions

Thusfar we assumed a frictionless market in evaluating mutual fund performance. It appears that the results are rather sensitive to the assumed initial set of benchmark assets. An investor who already owns a portfolio with efficient country weight allocation, and considers to extend this portfolio with an internationally investing mutual fund, can improve the risk-return tradeoff by taking a short position in most of the international mutual funds. However, an investor is typically confronted with short sales constraints on certain assets. Furthermore, when international investing is taken into account, the question whether an investor should directly invest in international assets or buy an internationally diversified mutual fund depends on the size of the transaction costs involved in buying the assets under consideration. Most of the mutual funds charge a load-fee for an investment into the fund. Moreover, some of the funds also have a back-end sales charge, although this percentage declines the longer the shares are held, and usually disappears entirely over time.

In considering the question whether a mean-variance optimizing investor can extend the efficient set by including additional assets into his portfolio, it is appropriate to incorporate market frictions. In case of short sales constraints, it is shown, for instance by Markowitz [1991] that the mean-variance frontier subject to short sales constraints consists of a finite number of segments of unrestricted frontiers. We denote the total number of segments by P. The assets in the efficient portfolios on the different P unrestricted frontiers coincide with the assets for which the short sales constraints are not binding in the optimization problem subject to short sales constraints (see, e.g. DeRoon, Nijman and Werker [1998]). Let $R_{t+1}^{(v)}$ denote such a L-dimensional subvector of R_{t+1} for which the short sales constraints in the restricted optimization problem are not binding. As shown by Luttmer [1996], if we include short sales constraints, the Law of One Price implies the following generalization of (3.1):

$$E[m_R(v)_{t+1} r_{t+1}] \leq \iota_N, \tag{3.4}$$

where the inequality sign reflects the short sales constraints on the additional assets. Since (3.4) holds, and a mean-variance stochastic discount factor is a linear function of the subset of L assets, the stochastic discount factor corresponding to mean-variance optimizing behavior that prices the assets on segment p of the restricted mean-variance frontier correctly is

$$m_R(v)_{t+1} = v + \alpha^{(v)\prime}(R_{t+1}^{(v)} - E[R_{t+1}^{(v)}]), \tag{3.5}$$

with

$$\alpha^{(v)} = Var[R_{t+1}^{(v)}]^{-1}(\iota_L - vE[R_{t+1}^{(v)}]).$$

Similar to the case without market frictions, a mean-variance investor cannot extend his efficient set of assets by including a mutual fund with gross return r_{t+1} if the stochastic discount factor given in (3.5) also prices r_{t+1} correctly, i.e. if it satisfies (3.4).

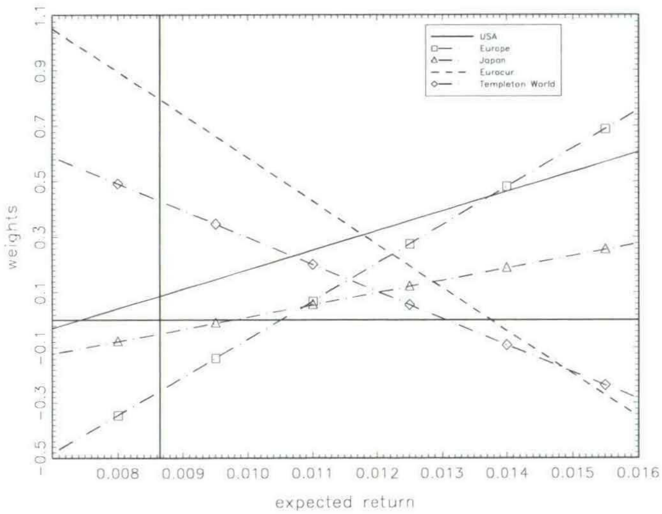
Since, in case of short sales restrictions, we can distinguish P different subsets of assets with corresponding mean-variance stochastic discount factors linear in those assets returns, the relationship between the Jensen measure and the performance measure defined in (3.3) now generalizes to

$$\lambda(v) = E[m_R(v)_{t+1} r_{t+1}] - 1 = v\alpha_J(v), \tag{3.6}$$

where $\alpha_J(v)$ is the generalized Jensen measure obtained as the intercept from a regression of a mutual fund's return in excess of the zero beta rate corresponding to the benchmark portfolios on a constant and the returns on the benchmark assets in subset p in excess of the same zero beta rate. The interpretation of performance measure (3.6) is that an investor with stochastic discount factor $m_R(v)_{t+1}$ can extend the efficient set by buying the mutual fund under consideration if and only if $\lambda(v) > 0$. In contrast, if $\lambda(v) \leq 0$ holds for one v, then the restricted mean-variance frontiers intersect, while if $\lambda(v) \leq 0$ holds for all v on all P subsets this corresponds with mean-variance spanning.

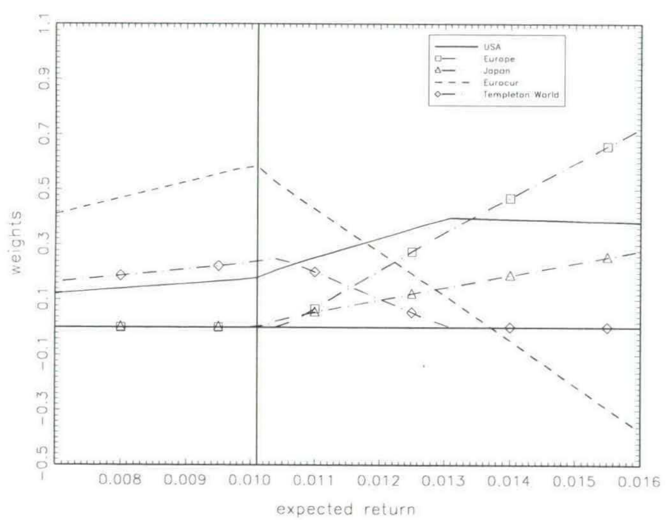
In order to illustrate that incorporating short sales restrictions seriously affects the diversification benefits of including mutual funds into the investor's portfolio, we show in Figure 3.1 the estimated unrestricted optimal portfolio weights for the initial set of four benchmark assets and the Templeton World mutual fund for a range of expected returns, assuming that all parameters coincide with their sample analogue as reported in Table 3.1.

Figure 3.1: Optimal Weights. The figure shows the optimal weights for four benchmark assets and the Templeton World mutual fund if short selling is not excluded. The horizontal line represents the zero weight line, while the vertical line represents the location of the Global Minimum Variance portfolio.



The global minimum variance portfolio corresponding to this set of initial assets is located at a monthly expected return of 0.87%. Note that the unrestricted efficient portfolio with an expected return of more than 1.31% (15.7% annually) contains a short position in the mutual fund. Moreover, a long position in the mutual fund coincides sometimes with short positions in a number of benchmark assets. Although this is not unrealistic for the currency factor, for the stock market indices a short position is not very realistic. In Figure 3.2 we show the corresponding optimal weights for the optimization problem under short sales restrictions for the Morgan Stanley USA, Europe and Japan indices, where we also imposed that an investor is no longer required to invest all his wealth in the available assets (see, e.g. Luttmer [1996]). This implies that an investor is allowed to take a long position in a riskless asset with zero return. The vertical line in the figure corresponds to the location of the portfolio where the investor does no longer take a position in the asset with zero return. It is straightforward to show that this portfolio corresponds to the tangency point of the line starting in the origin, i.e. expected return as well as variance equal to zero, to the restricted mean-variance frontier of the risky assets only. We will denote this portfolio as the zero-tangency portfolio.

Figure 3.2: Optimal Weights under short sales constraints. The figure shows the optimal weights for the four benchmark assets and the Templeton World fund under short sales constraints for the Morgan Stanley USA, Europe and Japan indices as well as Templeton World fund. The vertical line represents the location of the zero tangency portfolio, i.e. at the right hand side of this line investors invest all their wealth in the risky assets.



It appears that only investors whose optimal portfolio expected return lies below 1.31% will take a long position in the mutual fund. Note that since we did not impose short sales restrictions on the Eurocurrency portfolio, investors can still construct portfolios with very high expected returns. This would not be possible if we imposed short sales restrictions on all the assets. In contrast to the mean-variance optimization problem without restrictions, the optimal weights are piecewise linear in the expected portfolio return. As soon as a binding restriction for one of the assets is reached, a transition between two segments of the restricted mean-variance frontier occurs. In the sequel we will denote such a point as a transition point. Note that between two transition points the optimal weights for the assets behave linearly in the expected portfolio return.

In order to further illustrate the relevance of taking short sales restrictions into account, we also compute efficient frontiers for adding the other seventeen mutual funds to the benchmark assets, still assuming that all parameters coincide with the values reported in Table 3.1. In the case of two benchmark assets, a long position is taken for some values of the zero beta rate in thirteen out of eighteen mutual funds, while in the four benchmark case only eight mutual funds give a potential diversification benefit. However, for a restricted mean-variance efficient portfolio with an expected return of more than 1.70% the mutual funds will have zero weight in the four benchmark case. Recall that a zero optimal weight in a mutual fund can be caused by the short sales restriction on the benchmark assets as well as the short sales restriction on the mutual fund.

The short sales constraints assumed so far can be interpreted as an extreme transaction cost that investors have to pay for holding a short position. However, actual transaction costs may not completely prevent investors from taking short positions. Furthermore, there are also transaction costs associated with buying securities. When we incorporate transaction costs in analyzing performance of mutual funds, we have to take into account the investment horizon of the investor. By doing so, we can determine for what investment horizon a mutual fund gives a potential diversification benefit by including it in the investor's portfolio that outweighs transaction costs. Transaction costs can be handled by distinguishing between the return on a short and the return on a long position in the asset, as suggested in Luttmer [1996]. Under the assumption that returns are independently and identically distributed (i.i.d.), we construct a $2K$-dimensional vector $\tilde{R}_{t+1}$. The first K-elements contain the net returns on a long position in the benchmark assets, where a net return[4] is defined as $\tilde{R}^{l}_{i,t+1} = \tau^{l}_{i} R_{i,t+1}$, with $\tau^{l}_{i} = \frac{1}{1+a_i}$ where $a_i > 0$ are the transaction costs involved in taking a long position adjusted for the length of the investment horizon, i.e. $a_i = (1 + \tilde{a}_i)^{\frac{1}{H}} - 1$ where H is the investment horizon in months and $\tilde{a}_i$ is the cost per transaction. The second K-elements contain the net returns on a short position in the benchmark assets, defined as $\tilde{R}^{s}_{i,t+1} = \tau^{s}_{i} R_{i,t+1}$, with $\tau^{s}_{i} = \frac{1}{1-b_i}$ where $b_i > 0$ are the transaction costs involved in taking a short position adjusted for the length of the invest-

[4] Note that we define net returns as gross returns after transaction costs, while for instance Bekeart and Urias [1996] denote net returns as returns after operating expenses.

ment horizon[5]. In a similar way we construct a vector $\tilde{r}_{t+1}$ containing the long, $\tau^l r_{t+1}$, as well as short returns, $\tau^s r_{t+1,}$ on the mutual fund under consideration, i.e. after correcting for the load-fees that the fund charges.

Under the restriction that it is not possible to take a short position in the first K assets and a long position in the second K assets, we can, analogous to the case of short sales restrictions only, denote $\tilde{R}^{(v)}_{t+1}$ as a L-dimensional subvector of $\tilde{R}_{t+1}$ for which the constraints on the long and short position are not binding. Moreover, the stochastic discount factor linear in the subset of L assets that prices the subset of assets correctly is

$$\tilde{m}_R(v)_{t+1} = v + \tilde{\alpha}^{(v)\prime}(\tilde{R}^{(v)}_{t+1} - E[\tilde{R}^{(v)}_{t+1}]), \tag{3.7}$$

with

$$\tilde{\alpha}^{(v)} = Var[\tilde{R}^{(v)}_{t+1}]^{-1}(\iota_L - vE[\tilde{R}^{(v)}_{t+1}]).$$

It is now straightforward to show that, substituting the expressions of net long and short returns, condition (3.4) can be generalized to

$$\frac{1}{\tau^s_i} \leq E[\tilde{m}_R(v)_{t+1} r_{i,t+1}] \leq \frac{1}{\tau^l_i}, \quad i = 1..N, \tag{3.8}$$

where the inequality signs reflect the short and long sales constraints on the additional assets. In order to check whether an investor can extend the efficient set by including a mutual fund into his portfolio, we can determine

$$\tilde{\lambda}^{(p)}(v) = E[\tilde{m}_R(v)_{t+1} r_{t+1}] - 1 = v\tilde{\alpha}_J(v), \tag{3.9}$$

where $\tilde{\alpha}_J(v)$ is the generalized Jensen measure obtained as the intercept from a regression of a mutual fund's return in excess of the zero beta rate corresponding to the benchmark portfolios on a constant and the net returns on the benchmark assets in subset p in excess of the same zero beta rate. Now, an investor with stochastic discount factor $\tilde{m}_R(v)_{t+1}$ can extend the efficient set by buying the mutual fund under consideration if and only if $\tilde{\lambda}^{(p)}(v) > a_i$ or sell the mutual fund, if possible, if and only if $\tilde{\lambda}^{(p)}(v) < -b_i$. As may be clear, it is straightforward to adjust (3.8) and (3.9) when we impose short sales restrictions on a number of the benchmark assets as well as on the mutual funds under consideration. The empirical implications of incorporating transaction costs will become clear in the next section where we test for diversification benefits by including mutual funds into the initial portfolio.

[5] The transaction costs $a_i > 0$ and $b_i > 0$ can be interpreted as the ask and bid spread as a percentage of the price $P_{i,t}$ when buying or selling assets.

3.4 Testing for Spanning and Intersection in case of Market Frictions

The optimal weights reported in Figures 3.1 and 3.2 for the benchmark assets as well as the mutual fund are based on the point estimates of Table 3.1. Consequently, estimation errors in these parameters affect the possible diversification benefits by including the mutual funds into the portfolio. In order to test whether the diversification benefit that can be obtained from including a mutual fund[6] with return vector r_{t+1} to the initial set of benchmark assets with return vector R_{t+1} is significant, we want to test the hypothesis

$$E[m(v)_{t+1}r_{t+1}] = 1. \tag{3.10}$$

Recall that this corresponds to testing for a shift in the mean-variance efficient frontier. As is well known by now, this test can be based on the regression equation

$$r_{t+1} = a + BR_{t+1} + \varepsilon_{t+1}, \tag{3.11}$$

with $E[\varepsilon_{t+1}] = 0$ and $E[\varepsilon_{t+1}R_{t+1}] = 0$. Ignoring market frictions, spanning implies that $a = 0$ and $B\iota_K - 1 = 0$, and intersection of the extended mean-variance frontier and the mean-variance frontier of the original K assets implies that $av + (B\iota_K - 1) = 0$ for a given value v (Huberman and Kandel [1987], Bekaert and Urias [1996]). This test for mean-variance spanning can easily be extended for investors with other utility functions as shown by DeRoon, Nijman and Werker [1996]. Alternatively, GMM-tests can be used to test for intersection and spanning (see DeSantis [1994], Hansen, Heaton and Luttmer [1995] and Chen and Knez [1996]). Interpreting the outcome of this test in light of (3.3), spanning implies that no investors can extend their efficient set by taking a position in the mutual fund, while intersection means that only a particular group, i.e. the group with a stochastic discount factor with expectation v, cannot extend the efficient set of assets.

As shown by DeRoon, Nijman and Werker [1998], a test for a shift in the mean-variance frontier subject to short sales constraints can be implemented in a regression framework as well. Recall that the restricted mean-variance frontier consists of P segments of unrestricted mean-variance frontiers. This means that intersection between the initial and extended mean-variance frontiers can occur at P different unrestricted frontiers. Since the assets in the mean-variance portfolios on such a segment of the frontier coincide with the subset of L assets for which the short sales constraints are not binding in the restricted problem, and a stochastic

[6] The tests can easily be extended to the inclusion of a set of N mutual funds.

discount factor $m_R(v)_{t+1}$ is a linear function of the corresponding subvector $R_{t+1}^{(p)}$ only, implies that we can estimate the following P regressions

$$r_{t+1} = a^{(p)} + B^{(p)} R_{t+1}^{(p)} + \varepsilon_{t+1}^{(p)}, \tag{3.12}$$

and test whether

$$a^{(p)} v + (B^{(p)} \iota_L^{(p)} - 1) \leq 0 \tag{3.13}$$

holds for one value of v. If (3.13) holds for all v then the initial set of assets spans the extended set of assets. The inequality sign in (3.13) has to be replaced by an equality sign when there are no short sales restrictions on the additional assets.

Recall that a segment p of the restricted mean-variance frontier is bounded by two transition points. Since intersection at these two transition points implies spanning at segment p of the restricted mean-variance frontier, a test for mean-variance spanning is equivalent to testing whether (3.13) holds for two choices of v corresponding to these two transition points. We will denote the two values of v as: $v_{min}^{(p)}$ and $v_{max}^{(p)}$, the minimum and maximum expectation of the set of stochastic discount factors that price the subset of L assets correctly. The value of $v_{min}^{(p)}$ and $v_{max}^{(p)}$ can be determined as the inverse of the zero beta rates corresponding to the transition points bounding segment p of the restricted mean-variance frontier. Testing for spanning is therefore equivalent to testing whether the following two inequality restrictions

$$\begin{aligned} a^{(p)} v_{min}^{(p)} + (B^{(p)} \iota_L^{(p)} - 1) &\leq 0 \\ a^{(p)} v_{max}^{(p)} + (B^{(p)} \iota_L^{(p)} - 1) &\leq 0 \end{aligned} \tag{3.14}$$

hold jointly for $p = 1..P$. The joint one-sided inequality constraints can be tested with the Wald test under inequality constraints.

It is straightforward to show that a comparable regression framework as under short sales constraints only can be used for testing for a shift in the mean-variance frontier with transaction costs incorporated. Recall that $\tilde{R}_{t+1}^{(p)}$ denotes a L-dimensional subvector of the net return vector $\tilde{R}_{t+1}$ for which the constraints on the long and short position are not binding, then a test for mean-variance spanning when also transaction costs are incorporated, can be based upon whether in the P regressions

$$r_{t+1} = a^{(p)} + B^{(p)} \tilde{R}_{t+1}^{(p)} + \varepsilon_{t+1}^{(p)}, \tag{3.15}$$

the following restrictions hold jointly:

$$-b_i \leq a^{(p)} v_{min}^{(p)} + (B^{(p)} \iota_L^{(p)} - 1) \leq a_i \tag{3.16}$$

$$-b_i \ \leq \ a^{(p)}v^{(p)}_{max} + (B^{(p)}\iota^{(p)}_L - 1) \leq a_i$$

for $p = 1...P$, where a_i, b_i are the transaction costs for a long respectively short position in the fund under consideration, adjusted for the length of the investment horizon.

The joint two-sided constraints involved in mean-variance spanning with transaction costs can be tested using a Wald test. In case of a frictionless market, the Wald test statistic simply has a χ^2_{2N} distribution. However, in case of market frictions such as transaction costs and short selling constraints, we have to deal with inequality constraints. Analogous to the derivation of one-sided inequality constraints, as shown by, for instance, Kodde and Palm [1986], the Wald test statistic under two-sided inequality constraints:

$$\xi(v) = min_{-b_i \leq \alpha_J \leq a_i}(\hat{\alpha}_J(v) - \alpha_J(v))'Var(\hat{\alpha}_J(v))^{-1}(\hat{\alpha}_J(v) - \alpha_J(v)) \tag{3.17}$$

is asymptotically distributed as a mixture of χ^2 distributions, where the $2PN$ dimensional vector $\hat{\alpha}_J(v)$ corresponds to the left hand side of (3.14) and the $2PN \times 2PN$ covariance matrix $Var(\hat{\alpha}_J(v))$ can be obtained from the restricted covariance matrix of the OLS-estimates of (3.12).

In Tables 3.3 and 3.4 we present the outcomes of the test for mean-variance spanning in a frictionless market for the two respectively four benchmark case. Note that we no longer impose that an investor should invest all his wealth, implying that we assume that a risk free asset with zero return is available[7]. This assumption implies that the upper bound for v is 1. A lower bound for v can be obtained as the inverse of the intercept of the asymptote of the mean-variance frontier. It is straightforward to show that this intercept is equal to the expected return on the global minimum variance portfolio of the benchmark assets, which appear to be 0.940% and 0.886% for the two respectively four benchmark case.

In the two benchmark case it appears that the hypothesis of mean-variance spanning is rejected for eight mutual funds. This means that these funds can significantly extend a widely diversified international portfolio. When we extend the initial set of two assets to four benchmark assets, it appears that thirteen mutual funds give a significant extension of the investment set. Note that a potential diversification benefit can also mean that an investor has to take a short position in the mutual fund, i.e. a rejection of the spanning hypothesis can be caused by out as well as underperformance of the fund. To illustrate whether investors have to take short or long positions in the mutual funds for optimal diversification benefits, we computed for two values of v, one corresponding with a portfolio located near the zero tangency portfolio and one corresponding with a portfolio with an extreme high expected return, the performance measure

[7] It is not allowed to take a short position in a risk free asset with zero return.

Table 3.3: **Spanning and Intersection Tests in the case of two benchmark assets.** The table reports the interval of expected returns for which the hypothesis of intersection can be rejected as well as the p-values associated with the Wald tests for mean-variance spanning in a frictionless market. The initial set of benchmark assets consists of the Morgan Stanley World index and an equally weighted portfolio of Eurocurrency Deposits. Note that we do not impose that an investor should invest all his wealth. The (+) in the column labelled as 'z-t' indicates that an investor located near the zero tangency portfolio takes a long position in the fund for optimal diversification benefits, while a (+) in the column labelled 'asy' means that also for very high expected returns a long position in the fund is taken.

Two Benchmark Assets				
Mutual Fund	Rejection Interval Intersection in Expected Returns (%)	Spanning Test (frictionless) p-value	z-t	asy
Alliance Global Small.	$\langle 0.526, 0.969 \rangle$	0.000	(+)	(−)
Alliance Intl.	–	0.744	(−)	(−)
Bailard,Biehl Intl.	$\langle 0.933, 0.991 \rangle$	0.039	(−)	(−)
First Invest Global	–	0.305	(+)	(−)
Kemper Intl.	–	0.765	(−)	(−)
Lexington World Wide	$\langle 0.692, 0.983 \rangle$	0.000	(+)	(−)
New Perspective	$\langle 0.891, 1.025 \rangle$	0.000	(+)	(−)
Oppenheimer Global	–	0.810	(+)	(−)
Phoenix World Opport.	$\langle 0.719, 0.963 \rangle$	0.000	(−)	(−)
Pilot Kleinwort Intl.	–	0.566	(−)	(−)
Putnam Global Growth	$\langle 0.938, 0.956 \rangle$	0.121	(+)	(+)
Scudder Intl.	–	0.958	(+)	(+)
T. Rowe Price Intl.	–	0.283	(+)	(+)
Templeton Growth	$\langle 0.843, 1.040 \rangle$	0.000	(+)	(−)
Templeton Small Cmp.	$\langle 0.837, 1.040 \rangle$	0.000	(+)	(−)
Templeton World	$\langle 0.837, 1.037 \rangle$	0.000	(+)	(−)
United Intl. Growth	$\langle 0.930, 0.959 \rangle$	0.078	(+)	(+)
Vanguard Intl. Growth	$[0, 0.916\rangle$ and $\langle 0.939, +\infty)$	0.116	(+)	(+)

as given in (3.3). The (+)'s in the columns labelled 'z-t' in Tables 3.3 and 3.4 indicate that for investors whose portfolio is located near the zero tangency portfolio the fund shows outperformance and investors can extend the efficient set by taking a long position in the fund under consideration. Moreover, the (+)'s in the columns labelled 'asy' in Table 3.3 mean that also for very high expected returns, the funds still show outperformance and investors take a long position in the fund for optimal diversification benefits, while in the four benchmark case (Table 3.4) for very high expected returns only underperformance of the funds remains and investors take short positions in the mutual funds for diversification benefits.

The second column in the Tables 3.3 and 3.4 contain the intervals for the range of expected return values of the initial portfolio for which the hypothesis of intersection will be rejected in case there are no market frictions. For instance, in the four benchmark case, investors that initially hold a portfolio with an expected monthly return between 0.786% and 0.948% (i.e. between 9.4% and 11.4% annually) have a diversification benefit by including the Templeton World mutual fund. The rejection of the full spanning hypothesis is of course caused by the fact that the two frontiers differ substantially for this region of expected returns. If the corre-

Table 3.4: **Spanning and Intersection Tests in the case of four benchmark assets.** The table reports the interval of expected returns for which the hypothesis of intersection can be rejected as well as the p-values associated with the Wald test for mean-variance spanning in a frictionless market. The initial set of benchmark assets consists of the Morgan Stanley USA, Europe and Japan indices and an equally weighted portfolio of Eurocurrency Deposits. Note that we do not impose that an investor should invest all his wealth. The (+) in the column labelled 'z-t' indicates that an investor located near the zero tangency portfolio takes a long position in the fund for optimal diversification benefits, while a (+) in the column labelled 'asy' means that also for very high expected returns a long position in the fund is taken.

Four Benchmark Assets				
Mutual Fund	Rejection Interval Intersection in Expected Returns (%)	Spanning Test (frictionless) p-value	z-t	asy
Alliance Global Small.	$[0,0.811\rangle$ and $\langle 0.932,+\infty\rangle$	0.067	(−)	(−)
Alliance Intl.	$[0,0.829\rangle$ and $\langle 0.879,+\infty\rangle$	0.117	(−)	(−)
Bailard,Biehl Intl.	$\langle 0.895, 1.168\rangle$	0.003	(−)	(−)
First Invest Global	$\langle 0.795, 0.897\rangle$	0.039	(−)	(−)
Kemper Intl.	$\langle 0.552, 0.923\rangle$	0.000	(+)	(−)
Lexington World Wide	$[0,0.447\rangle$ and $\langle 0.886,+\infty\rangle$	0.035	(−)	(−)
New Perspective	$\langle 0.843, 0.924\rangle$	0.024	(+)	(−)
Oppenheimer Global	–	0.819	(−)	(−)
Phoenix World Opport.	$[0,0.961\rangle$ and $\langle 1.531,+\infty\rangle$	0.146	(−)	(−)
Pilot Kleinwort Intl.	$\langle 0.902, 1.632\rangle$	0.002	(−)	(−)
Putnam Global Growth	$\langle 0.857, 0.982\rangle$	0.117	(+)	(−)
Scudder Intl.	$\langle 0.761, 0.911\rangle$	0.007	(+)	(−)
T. Rowe Price Intl.	$\langle 0.827, 0.911\rangle$	0.029	(+)	(−)
Templeton Growth	$\langle 0.786.0.957\rangle$	0.000	(+)	(−)
Templeton Small Cmp.	$\langle 0.779, 0.966\rangle$	0.000	(+)	(−)
Templeton World	$\langle 0.786, 0.948\rangle$	0.000	(+)	(−)
United Intl. Growth	$\langle 0.800, 0.932\rangle$	0.002	(+)	(−)
Vanguard Intl. Growth	$\langle 0.843, 0.918\rangle$	0.032	(+)	(−)

sponding entry in the column is empty, intersection cannot be rejected for any choice of the expected return.

In order to analyze the impact of frictions we first of all present the mean-variance spanning tests imposing transaction costs. In order to test the hypothesis whether it is efficient to invest directly in international assets or to buy an internationally diversified mutual fund, we also assume that there are transaction costs involved in taking a position in the benchmark assets. Following Luttmer [1996] we impose transaction costs for the benchmark assets that equal 0.5% for buying as well as selling. In Table 3.5 we present for the case of two benchmark assets, the transaction costs a mutual fund may charge such that the hypothesis of mean-variance spanning is just rejected at the 5% level for different investment horizons. It appears that for an investment horizon of only one month all the mutual funds in the sample give a diversification benefit in the case of two benchmark assets when we incorporate transaction costs. This can probably be explained by the fact that we fixed the transaction costs for the benchmark assets at 0.5% for buying as well as selling, which makes these assets relatively expensive compared to the mutual funds. For holding periods of more than six months only seven mutual funds

Table 3.5: **Spanning Tests imposing transaction costs.** The table shows the transaction costs a fund may charge such that the hypothesis of mean-variance spanning is just rejected at the 5% level for various investment horizons. The initial benchmark assets are the Morgan Stanley World index and an equally weighted portfolio of Eurocurrency Deposits.

Mutual Fund	Holding Period (in months) 1	6	12	18	24	30	36
Alliance Global Small.	0.40	0.95	2.00	3.10	4.15	5.30	>6.00
Alliance Intl.	0.40	-	-	-	-	-	-
Bailard,Biehl Intl.	0.15	-	-	-	-	-	-
First Invest Global	0.35	-	-	-	-	-	-
Kemper Intl.	0.30	-	-	-	-	-	-
Lexington World Wide	0.55	1.15	2.45	3.75	5.00	>6.00	>6.00
New Perspective	0.80	0.90	1.30	1.90	2.45	3.10	3.75
Oppenheimer Global	0.45	-	-	-	-	-	-
Phoenix World Opport.	0.20	0.65	1.45	2.25	3.05	3.90	4.70
Pilot Kleinwort Intl.	0.30	-	-	-	-	-	-
Putnam Global Growth	0.70	-	-	-	-	-	-
Scudder Intl.	0.45	-	-	-	-	-	-
T. Rowe Price Intl.	0.60	-	-	-	-	-	-
Templeton Growth	0.90	1.55	2.65	3.90	5.20	>6.00	>6.00
Templeton Small Cmp.	1.00	1.70	3.15	4.70	>6.00	>6.00	>6.00
Templeton World	0.95	1.45	2.60	3.90	5.20	>6.00	>6.00
United Intl. Growth	0.60	-	-	-	-	-	-
Vanguard Intl. Growth	0.60	-	-	-	-	-	-

give a significant improvement in the risk-return trade-off. An empty entry in a column indicates that the fund does not provide any diversification benefits, even in case of zero transaction costs. Comparing the outcomes of Table 3.5 to the actual load-fees that the funds charge, it appears that an investment horizon (holding period) of about two years is required to have a significant improvement in the risk-return trade-off.. An exception is Lexington World Wide, which can be marked as a no-load fund, that gives diversification benefits for all the holding periods by including it in the investor's portfolio.

In Table 3.6 we present the outcomes of the test for mean-variance spanning imposing short sales restrictions on the Morgan Stanley World, USA, Europe and Japan indices as well as the mutual fund under consideration. We do not impose a short sales restriction on the Eurocurrency Deposits portfolio.

It appears that in the case of two benchmark assets, the hypothesis of mean-variance spanning under short sales restrictions is rejected for four mutual funds, indicating that these funds still show outperformance. Consequently, investors who own a portfolio with predetermined country weight allocation can improve their portfolio's risk-return trade-off by including one of these mutual funds. However, in the case of four benchmark assets the hypothesis of mean-variance spanning is not rejected anymore. Apparently, the combination of short sales restrictions on the benchmark assets as well as on the mutual funds makes the diversification benefit

Table 3.6: **Spanning Tests under short sales constraints.** The table reports the Wald test statistic under inequality constraints and the corresponding p-value for mean-variance spanning. Note that we do not impose short sales restrictions on the currency hedge portfolio in both sets of benchmark assets.

Mutual Fund	Two Benchmark Assets p-value	Four Benchmark Assets p-value
Alliance Global Small.	0.522	0.747
Alliance Intl.	0.912	0.817
Bailard,Biehl Intl.	0.713	0.802
First Invest Global	0.552	0.845
Kemper Intl.	1.000	0.734
Lexington World Wide	0.228	0.792
New Perspective	0.017	0.487
Oppenheimer Global	0.435	0.923
Phoenix World Opport.	0.567	0.744
Pilot Kleinwort Intl.	0.773	0.725
Putnam Global Growth	0.164	0.743
Scudder Intl.	0.505	0.806
T. Rowe Price Intl.	0.299	0.671
Templeton Growth	0.002	0.291
Templeton Small Cmp.	0.006	0.189
Templeton World	0.002	0.368
United Intl. Growth	0.187	0.500
Vanguard Intl. Growth	0.228	0.495

that appeared to be present in the frictionless market disappear completely. Moreover, the outperformance present in the two benchmark case is for portfolios with efficient country weight allocation not present anymore. So, although Figure 3.2 suggested that a mean-variance optimizing investor takes a long position in the Templeton World fund, it appears that there is not a significant diversification benefit by taking a long position in the mutual fund after imposing short sales constraints.

3.5 Conditional Performance Evaluation

Recent studies show that returns on stocks and bonds are predictable over time (see e.g., Ferson and Harvey [1993], Keim and Stambaugh [1986]). Time-varying expected returns and (co)variances imply time-varying mean-variance frontiers. Consequently, mean-variance optimizing investors will dynamically adjust their portfolios because of the changing economic conditions. Therefore it can be the case that under certain economic conditions there are diversification benefits for a mean-variance optimizing investor by including additional assets into his portfolio while under different circumstances these benefits are absent. The implication for performance evaluation of mutual funds is, as advocated in recent papers of Ferson and

Schadt [1996] and Chen and Knez [1996], that the evaluation should be conditional upon the state of the economy. The aim of conditional performance evaluation is to distinguish mutual funds with real timing or selection ability of the fund manager from managed portfolio strategies that can be replicated using publicly available information.

In previous sections we assumed that the expected returns on the assets and the corresponding (co)variances are constant over time. We will now relax this assumption. Let us first of all consider the case of a frictionless market in conditional performance evaluation of mutual funds. Using a set of information variables z_t, supposed to reflect the state of the economy, a test for conditional mean-variance spanning can be based on the following regression:

$$r_{t+1} = a + \gamma' z_t + BR_{t+1} + \varepsilon_{t+1}. \tag{3.18}$$

One can easily test for mean-variance spanning for arbitrary values of the information variables z_t as well as for mean-variance spanning for specific values of z_t (see Appendix 3.A for further details). The first case coincides with $a = \gamma = 0$ and $B\iota_K - 1 = 0$, while the second case holds for $a = -\gamma' z_t$ and $B\iota_K - 1 = 0$. Alternatively, one can incorporate conditional information by adding so-called scaled returns to the regression equation (3.18) (see, e.g. Cochrane [1997], Bekeart and Urias [1996]). However, the disadvantage of this method of conditional performance evaluation is the dimensionality problem that arises in estimating and testing when the set of information variables or the set of initial benchmark assets is large.

Following previous studies on conditional performance evaluation (see, Ferson and Schadt [1996] and Chen and Knez [1996]) we use the following set of information variables: 1) the lagged level of the one-month Tbill yield, 2) the lagged dividend yield on the Morgan Stanley World index (in the two benchmark case), 3) the lagged term spread measured as the difference between a constant maturity 10-year bond yield and a constant maturity 1-year bond yield and 4) a dummy for the month of January.

Since rejection of the hypothesis of mean-variance spanning for arbitrary values of the information variables z_t does not imply that mean-variance spanning is absent under specific economic circumstances, we consider only tests for the hypothesis of mean-variance spanning under a number of sets of specific values for the information variables lagged Tbill yield, lagged dividend yield and lagged term spread. Moreover, we examine whether it affects the diversification benefits when these specific economic circumstances occur in the month January or in the other months of the year. In Table 3.7 we report outcomes for tests of the hypothesis of mean-variance spanning in a frictionless market, conditional upon three different sets of information variables, when the initial set of assets consist of the Morgan Stanley World and the currency hedge portfolio, i.e. the two benchmark case.

Table 3.7: **Conditional Spanning Tests in a frictionless market.** The table reports outcomes of a test for conditional mean-variance spanning ignoring market frictions, in the case of two benchmark assets, for specific values of the information variables. Panel A of the table reports a number of sets of specific values for the information variables and their average value (between parentheses) over the sample period 1982-1994. Panel B reports the p-values associated with the Wald test for mean-variance spanning in a frictionless market conditonal upon the values for the information variables.

Panel A: specific values information variables (annualized)						
Variable (average)	set 1		set 2		set 3	
Term spread (1.6%)	0.4%	0.4%	1.6%	1.6%	1.6%	1.6%
Month	January	Other	January	Other	January	Other
Tbill yield (6.0%)	6.0%	6.0%	6.0%	6.0%	2.4%	2.4%
Div yield (3.1%)	3.1%	3.1%	3.1%	3.1%	3.1%	3.1%
Panel B: p-values mean-variance spanning tests in frictionless market						
Mutual Fund						
Alliance Global Small.	0.000	0.000	0.000	0.000	0.000	0.000
Alliance Intl.	0.431	0.813	0.367	0.830	0.429	0.767
Bailard,Biehl Intl.	0.009	0.030	0.014	0.038	0.007	0.021
First Invest Global	0.010	0.208	0.034	0.277	0.070	0.370
Kemper Intl.	0.274	0.648	0.131	0.651	0.376	0.652
Lexington World Wide	0.000	0.000	0.000	0.000	0.000	0.000
New Perspective	0.000	0.000	0.000	0.000	0.000	0.000
Oppenheimer Global	0.132	0.015	0.670	0.105	0.105	0.031
Phoenix World Opport.	0.000	0.000	0.000	0.000	0.000	0.000
Pilot Kleinwort Intl.	0.821	0.787	0.814	0.590	0.815	0.794
Putnam Global Growth	0.107	0.119	0.101	0.098	0.115	0.118
Scudder Intl.	0.957	0.942	0.867	0.960	0.954	0.981
T. Rowe Price Intl.	0.383	0.246	0.186	0.224	0.426	0.317
Templeton Growth	0.000	0.000	0.000	0.000	0.000	0.000
Templeton Small Cmp.	0.000	0.000	0.000	0.000	0.000	0.000
Templeton World	0.000	0.000	0.000	0.000	0.000	0.000
United Intl. Growth	0.051	0.015	0.063	0.022	0.023	0.004
Vanguard Intl. Growth	0.160	0.186	0.119	0.112	0.186	0.194

Table 3.7 indicates that under specific economic circumstances some mutual funds give diversification benefits while these are absent in other circumstances. Moreover, some mutual funds only provide investors with an improved risk-return trade-off in January while in the rest of the year the fund does not give any diversification benefits. For instance, First Invest Global gives diversification benefits in January conditional upon information set 1 and 2, i.e. the term spread equal to or lower than the average value and the Tbill yield and Dividend yield equal to the average value over the period 1982-1994. However, when the Tbill yield is below the average over the sample period (set 3) then the fund does not give an improvement in the risk-return trade-off in any of the months. If we compare the outcomes of the conditional mean-variance spanning test with the unconditional mean-variance spanning test (Table 3.3) then it appears that, not surprisingly, outperformance is found roughly for the same funds but that for some funds (First Invest Global, Oppenheimer Global) distinction can be made when the fund under- or outperforms.

Similar to the unconditional mean-variance spanning tests, conditional mean-variance spanning tests with market frictions incorporated can be based upon a regression framework (see Appendix 3.A for further details). In case of transaction costs, a test for conditional mean-variance spanning can now be based on testing whether in the P regressions

$$r_{t+1} = a^{(p)} + \gamma^{(p)\prime} z_t + B^{(p)} \tilde{R}^{(p)}_{t+1} + \varepsilon^{(p)}_{t+1} \tag{3.19}$$

the following restrictions hold jointly:

$$\begin{aligned} -b_i &\leq a^{(p)} v^{(p)}_{min} + \gamma^{(p)\prime} \tilde{z} v^{(p)}_{min} + (B^{(p)} \iota^{(p)}_L - 1) \leq a_i \\ -b_i &\leq a^{(p)} v^{(p)}_{max} + \gamma^{(p)\prime} \tilde{z} v^{(p)}_{max} + (B^{(p)} \iota^{(p)}_L - 1) \leq a_i \end{aligned} \tag{3.20}$$

for $p = 1...P$, where $\tilde{z}$ denotes a specific choice for the information variables, $\tilde{R}^{(p)}_{t+1}$ is, as before, the L-dimensional net return vector of the initial assets for which the constraints on the long and short position are not binding, $v^{(p)}_{min}$ and $v^{(p)}_{max}$ are the inverses of the zero beta rates corresponding to transition points bounding segment p of the conditional mean-variance frontier and a_i, b_i are the transaction costs involved in taking a long respectively short position. The joint constraints in (3.20) can be tested with the Wald test under two-sided inequality constraints given in (3.17).

Comparable with the unconditional case, we impose transaction costs for the benchmark assets that equal 0.5% for buying as well as selling. Taking a position in the currency hedge portfolio is assumed to be free of charge. In contrast to the unconditional case, we now fix the total transaction costs involved in taking a long or short position in the mutual fund at 0.5%, and we test whether under specific economic circumstances, $\tilde{z}$, a mutual fund provides diversification benefits in the case of two benchmark assets. Note that we consider only an investment horizon (holding period) of one month. Table 3.8 presents the outcomes for the conditional mean-variance spanning tests with transaction costs incorporated.

It appears that in case of transaction costs the potential diversification benefits are rather sensitive for the specific values of the information variables $\tilde{z}$. For instance, when the term spread, Tbill yield and dividend yield are almost equal to their average value over the sample period (set 2), only two mutual funds give an improvement in the risk-return trade-off in January. However, in the other months of the year, ten mutual funds show a significant shift in the conditional mean-variance frontier. A similar pattern is observed for other values of the information variables $\tilde{z}$, suggesting that a dynamic trading strategy of taking a position in the mutual funds under consideration in eleven months of the year, and not having a position in the mutual funds in January, gives optimal diversification benefits.

Table 3.8: **Conditional Spanning Tests imposing transaction costs.** The table reports outcomes of a test for conditional mean-variance spanning where transaction costs of 0.5 are taken into account, in the case of two benchmark assets, for specific values of the information variables. Panel A of the table reports a number of sets of specific values for the information variables and their average value (between parentheses) over the sample period 1982-1994. Panel B reports the p-values associated with the Wald test under two-sided inequality constraints for mean-variance spanning conditional upon the corresponding set of the information variables.

Panel A: specific values information variables (annualized)						
Variable (average)	set 1		set 2		set 3	
Term spread (1.6%)	0.4%	0.4%	1.6%	1.6%	1.6%	1.6%
Month	January	Other	January	Other	January	Other
Tbill yield (6.0%)	6.0%	6.0%	6.0%	6.0%	2.4%	2.4%
Div yield (3.1%)	3.1%	3.1%	3.1%	3.1%	3.1%	3.1%
Panel B: p-values mean-variance spanning tests with transaction costs						
Mutual Fund						
Alliance Global Small.	0.257	0.154	0.367	0.055	0.193	0.155
Alliance Intl.	1.000	0.724	1.000	0.142	0.990	0.943
Bailard,Biehl Intl.	0.632	0.997	0.943	0.748	0.394	0.857
First Invest Global	0.002	0.106	0.011	0.373	0.049	0.527
Kemper Intl.	1.000	0.461	1.000	0.088	1.000	0.587
Lexington World Wide	0.284	0.018	0.612	0.002	0.105	0.014
New Perspective	0.153	0.006	0.426	0.000	0.219	0.041
Oppenheimer Global	0.047	0.000	0.425	0.006	0.043	0.003
Phoenix World Opport.	0.417	0.423	0.377	0.114	0.147	0.119
Pilot Kleinwort Intl.	0.887	0.790	0.841	0.372	0.941	0.906
Putnam Global Growth	0.969	0.278	0.948	0.002	0.923	0.396
Scudder Intl.	0.908	0.416	0.968	0.069	0.953	0.624
T. Rowe Price Intl.	1.000	0.089	1.000	0.002	0.999	0.218
Templeton Growth	0.276	0.037	0.313	0.000	0.214	0.054
Templeton Small Cmp.	0.009	0.002	0.034	0.000	0.003	0.002
Templeton World	0.315	0.042	0.335	0.000	0.198	0.046
United Intl. Growth	0.374	0.012	0.715	0.001	0.110	0.005
Vanguard Intl. Growth	1.000	0.301	1.000	0.008	0.999	0.423

The final step is to incorporate short sales constraints on certain assets. Recall that short sales constraints can be interpreted as extreme transaction costs that investors have to pay for taking a short position in the assets under consideration. Therefore we can, similar to the unconditional mean-variance spanning tests with short sales restrictions, estimate the following P regressions

$$r_{t+1} = a^{(p)} + \gamma^{(p)\prime} z_t + B^{(p)} R^{(p)}_{t+1} + \varepsilon^{(p)}_{t+1} \tag{3.21}$$

and test whether the following two inequality restrictions

$$\begin{aligned} a^{(p)} v^{(p)}_{min} + \gamma^{(p)\prime} \tilde{z} v^{(p)}_{min} + (B^{(p)} \iota^{(p)}_L - 1) &\leq 0 \\ a^{(p)} v^{(p)}_{max} + \gamma^{(p)\prime} \tilde{z} v^{(p)}_{max} + (B^{(p)} \iota^{(p)}_L - 1) &\leq 0 \end{aligned} \tag{3.22}$$

hold jointly for $p = 1...P$. The interpretation of $v^{(p)}_{min}$ and $v^{(p)}_{max}$ is similar to the one in (3.20). Note that the inequality sign in (3.22) has to be replaced by an equality sign when there are no short sales constraints on the additional assets. The joint constraints (3.22) can be tested using the Wald test under one-sided inequality constraints (see, for instance, DeRoon, Nijman and Werker [1998]). Table 3.9 presents the outcomes for the conditional mean-variance spanning tests when we impose short sales constraints on the Morgan Stanley World index and the fund under consideration.

Table 3.9: **Conditional Spanning Tests with short sales restrictions.** The table reports outcomes of a test for restricted conditional mean-variance spanning, in the case of two benchmark assets, for specific values of the information variables. Note that we do not impose short sales restrictions on the currency hedge portfolio. Panel A of the table reports a number of sets of specific values for the information variables and their average value (between parentheses) over the sample period 1982-1994. Panel B reports the p-values associated with the Wald test under one-sided inequality constraints on mean-variance spanning conditional upon the corresponding set of the information variables.

Panel A: specific values information variables (annualized)						
Variable (average)	set 1		set 2		set 3	
Term spread (1.6%)	0.4%	0.4%	1.6%	1.6%	1.6%	1.6%
Month	January	Other	January	Other	January	Other
Tbill yield (6.0%)	6.0%	6.0%	6.0%	6.0%	2.4%	2.4%
Div yield (3.1%)	3.1%	3.1%	3.1%	3.1%	3.1%	3.1%
Panel B: p-values mean-variance spanning tests with short sales constraints						
Mutual Fund						
Alliance Global Small.	0.165	0.146	0.259	0.312	0.108	0.119
Alliance Intl.	0.580	0.995	0.637	0.903	0.623	0.615
Bailard,Biehl Intl.	0.516	0.572	0.542	0.576	0.510	0.515
First Invest Global	0.002	0.124	0.016	0.773	0.033	0.357
Kemper Intl.	0.618	0.386	0.553	0.358	0.653	0.380
Lexington World Wide	0.153	0.024	0.403	0.041	0.052	0.007
New Perspective	0.120	0.010	0.320	0.006	0.142	0.037
Oppenheimer Global	0.039	0.002	0.317	0.095	0.023	0.004
Phoenix World Opport.	0.245	0.320	0.255	0.320	0.065	0.077
Pilot Kleinwort Intl.	0.880	0.646	0.873	0.769	0.882	0.953
Putnam Global Growth	0.828	0.344	0.744	0.135	0.851	0.312
Scudder Intl.	0.899	0.378	0.805	0.444	0.893	0.439
T. Rowe Price Intl.	0.812	0.132	0.596	0.120	0.773	0.200
Templeton Growth	0.164	0.040	0.224	0.001	0.122	0.040
Templeton Small Cmp.	0.008	0.003	0.024	0.001	0.004	0.001
Templeton World	0.188	0.058	0.230	0.003	0.114	0.038
United Intl. Growth	0.238	0.024	0.970	0.021	0.062	0.006
Vanguard Intl. Growth	0.696	0.272	0.596	0.155	0.762	0.318

It appears that, similar to the case of conditional mean-variance spanning with transaction costs incorporated, mutual funds provide less diversification benefits in January compared to the rest of the year. Recall that the diversification benefits present in the conditional mean-variance spanning with transaction costs incorporated can be due to out as well as underperformance of the fund. Moreover, to have an improvement in the risk-return trade-off of his initial portfolio it can also mean that the investor has to take a short position in some of the benchmark assets. The combination of short sales restrictions on the mutual funds as well as on some of the benchmark assets therefore leads to less diversification benefits of the mutual funds compared to the case of transaction costs only.

3.6 Concluding Remarks

In this chapter we examined whether internationally investing U.S. based mutual funds can extend an investor's efficient investment set. It appears that the answer to this question depends first of all on the assumed set of benchmark assets supposed to reflect the current portfolio choice, and secondly, on the assumption of a frictionless market. Using simple linear regressions we tested for mean-variance spanning in a frictionless as well as a market with transaction costs and short sales restrictions incorporated. Furthermore, it has been shown that these tests for mean-variance spanning are closely related to the issue of performance evaluation of mutual funds.

A risk-averse mean-variance optimizing investor that initially holds a widely diversified international portfolio with predetermined country weight allocation and considers to extend his portfolio with an internationally diversified mutual fund can improve his portfolio risk-return trade-off by taking long or short positions in the mutual funds. Although transaction costs and short sales constraints reduce the set of mutual funds that give potential diversification benefits, an extension of the portfolio with an internationally investing mutual fund is still worthwhile. Alternatively, this can be interpreted that even after imposing transaction costs and short sales constraints some mutual funds still show outperformance for investors with an investment strategy with predetermined country weight allocation. However, most mean-variance optimizing investors that already own an internationally diversified portfolio with efficient country weight allocation can only have an improvement in the risk-return trade-off by taking short positions in the mutual funds. Consequently, incorporating short sales restrictions seriously affects these potential diversification benefits for this group of the investing public. Moreover, it means that internationally diversified mutual funds do not show outperformance for investors that already own a diversified portfolio of international stocks with efficient country weight allocation.

Appendix 3.A

Testing for Conditional Mean-Variance Spanning

In this appendix we show how to test for mean-variance spanning in a frictionless as well as a market with short sales constraints and transaction costs when we incorporate conditional information. Recall that in a frictionless market there exists a stochastic discount factor M_{t+1} such that

$$E[M_{t+1}R_{t+1}|\, I_t] = \iota_K, \tag{A.1}$$

where ι_K is a K-vector of ones and I_t is the public information set available at time t. Denote z_t as a set of information variables, including a constant, supposed to reflect the state of the economy, and assume that

$$\begin{aligned} E[R_{t+1}|z_t] &= \gamma_R' z_t, \\ E[r_{t+1}|z_t] &= \gamma_r' z_t. \end{aligned}$$

The mean-variance stochastic discount factor $m(v)_{t+1}$ given by

$$m(v)_{t+1} = v + \alpha(v)'(R_{t+1} - E[R_{t+1}|z_t]), \tag{A.2}$$

with

$$\alpha(v) = Var[R_{t+1}|z_t]^{-1}(\iota_K - vE[R_{t+1}|z_t])$$

has the lowest variance of all stochastic discount factors with expectation v that price R_{t+1}correctly. Denote $Var[R_{t+1}|z_t]$ as Σ_{RR} and $Cov[r_{t+1}, R_{t+1}|z_t]$ as Σ_{rR}. Now, a mean-variance optimizing investor will not have a diversification benefit if the stochastic discount factor given in (A.2) also prices the mutual fund's return r_{t+1} correctly. This implies that

$$\begin{aligned} E_t[m(v)_{t+1}r_{t+1}] &= 1 \Leftrightarrow \\ v\gamma_r' z_t + \Sigma_{rR}\Sigma_{RR}^{-1}(\iota_K - v\gamma_R' z_t) &= 1 \Leftrightarrow \\ (\gamma_r' - \Sigma_{rR}\Sigma_{RR}^{-1}\gamma_R')vz_t + (\Sigma_{rR}\Sigma_{RR}^{-1}\iota_K - 1) &= 0. \end{aligned} \tag{A.3}$$

If this equality holds for one value of v then there is intersection, while if (A.3) holds for all v then there is spanning. It is now straightforward to show that $(\gamma_r' - \Sigma_{rR}\Sigma_{RR}^{-1}\gamma_R')$ and $\Sigma_{rR}\Sigma_{RR}^{-1}$ can consistently be estimated by the OLS estimates for γ and B in the following regression equation

$$r_{t+1} = \gamma' z_t + BR_{t+1} + \varepsilon_{t+1}, \tag{A.4}$$

with $E[\varepsilon_{t+1} z_t] = E[\varepsilon_{t+1} R_{t+1}] = 0$. Consequently, the hypothesis that there is intersection for a given value of v and z_t can be tested by testing

$$\gamma' \tilde{z} v + (B\iota_K - 1) = 0, \tag{A.5}$$

and the hypothesis of spanning for arbitrary values of z_t can be tested by testing

$$\gamma = 0 \text{ and } (B\iota_K - 1) = 0, \tag{A.6}$$

while spanning for specific values of z_t occurs for

$$\gamma' \tilde{z} = 0 \text{ and } (B\iota_K - 1) = 0, \tag{A.7}$$

where $\tilde{z}$ denotes a specific choice for the information variables.

When we incorporate short sales constraints (A.1) generalizes to

$$E_t[m_R(v)_{t+1} r_{t+1}] \leq 1, \tag{A.8}$$

where $m_R(v)_{t+1}$ is the stochastic discount factor that prices the subset of L assets on segment p of the restricted mean-variance frontier correctly. This implies that (A.3) generalizes to

$$(\gamma_r' - \Sigma_{rR}^{(p)} \Sigma_{RR}^{-1(p)} \gamma_R') v z_t + (\Sigma_{rR}^{(p)} \Sigma_{RR}^{-1(p)} \iota_K - 1) \leq 0. \tag{A.9}$$

As before, it is straightforward to show that $(\gamma_r' - \Sigma_{rR}^{(p)} \Sigma_{RR}^{-1(p)} \gamma_R')$ and $\Sigma_{rR}^{(p)} \Sigma_{RR}^{-1(p)}$ can consistently be estimated by the OLS estimates for $\gamma^{(p)}$ and $B^{(p)}$ in the regression equation

$$r_{t+1} = \gamma^{(p)\prime} z_t + B^{(p)} R_{t+1}^{(p)} + \varepsilon_{t+1}. \tag{A.10}$$

Recall that the restricted mean-variance frontier consists of P segments. Denote $v_{min}^{(p)}$ and $v_{max}^{(p)}$ as the minimum and maximum expectation of the set of stochastic discount factors that price the subset of L assets correctly, then testing for spanning for specific values of z_t is therefore equivalent to testing whether the following system of inequality restrictions

$$\begin{aligned} \gamma^{(p)\prime} \tilde{z} v_{min}^{(p)} + (B^{(p)} \iota_L^{(p)} - 1) &\leq 0 \\ \gamma^{(p)\prime} \tilde{z} v_{max}^{(p)} + (B^{(p)} \iota_L^{(p)} - 1) &\leq 0 \end{aligned} \tag{A.11}$$

hold jointly for $p = 1..P$.

Comparable with the unconditional case, transaction costs can be handled by distinguishing between the return on a short and a long position in the assets. Denote $\tau_i^l = \frac{1}{1+a_i}$ and $\tau_i^s = \frac{1}{1-b_i}$

as the transaction costs involved in taking a long respectively short position in the assets, and construct a $2K$-dimensional vector $\tilde{R}_{t+1}$ where the first K elements contain the net returns on a long position and the second K elements the net returns on a short position in the benchmark assets. Now it is straightforward to show that condition (A.8) can be generalized to

$$\frac{1}{\tau^s} \leq E_t[\tilde{m}_R(v)_{t+1} r_{t+1}] \leq \frac{1}{\tau^l}, \tag{A.12}$$

where $\tilde{m}_R(v)_{t+1}$ is the stochastic discount factor that prices the subset of L assets correctly, and the inequality signs reflect the short and long sales constraints on the additional asset. Substituting the expression for $\tilde{m}_R(v)_{t+1}$ in (A.12) gives

$$-b_i \leq (\gamma_r' - \Sigma_{rR}^{(p)} \Sigma_{RR}^{-1(p)} \gamma_R') v z_t + (\Sigma_{rR}^{(p)} \Sigma_{RR}^{-1(p)} \iota_K - 1) \leq a_i, \tag{A.13}$$

then a test for conditional mean-variance spanning for specific values of z_t when also transaction costs are incorporated can be based upon whether in the P regressions

$$r_{t+1} = \gamma^{(p)\prime} z_t + B^{(p)} \tilde{R}_{t+1}^{(p)} + \varepsilon_{t+1}, \tag{A.14}$$

the following restrictions hold jointly

$$\begin{aligned} -b_i &\leq \gamma^{(p)\prime} \tilde{z} v_{min}^{(p)} + (B^{(p)} \iota_L^{(p)} - 1) \leq a_i \\ -b_i &\leq \gamma^{(p)\prime} \tilde{z} v_{max}^{(p)} + (B^{(p)} \iota_L^{(p)} - 1) \leq a_i. \end{aligned} \tag{A.15}$$

Chapter 4

Style Analysis and Performance Evaluation of Dutch Mutual Funds

In this chapter we show how style analysis of mutual funds can be used to circumvent the problem of self-reported investment styles, and to improve relative performance evaluation. Subsequently, we relate style analysis to performance evaluation and present results on the performance of Dutch mutual funds. Most strikingly, Dutch mutual funds that mainly invest in Netherlands equity show relative outperformance of the passive portfolio of indices reflecting the mutual fund's investment style. Moreover, the same group of funds provide an extension of the mean-variance efficient investment set for Dutch investors, even after taking short sales restrictions into account, indicating that a domestic market effect might be present.

4.1 Introduction

Differences in exposure to investment styles can explain a large part of the cross-sectional variation in mutual fund returns. Nevertheless, many investors and the financial press often simply compare realized mutual fund returns without taking differences in exposures into account. In so-called relative performance evaluation, mutual fund returns are compared with each other or with a benchmark asset covering the fund's investment style. Mutual fund managers that are aware of this fact can improve the outcome of relative performance evaluation by investing in securities that are not in accordance with their stated investment style or objective (see, Brown and Goetzmann [1997]).

In order to avoid gaming of benchmark assets, return-based style analysis, introduced by Sharpe [1992], can be used as an objective instrument to determine the mutual fund's investment style. After having determined the effective investment mix of a portfolio, performance evaluation can be based simply on a comparison of the mutual fund's return with a similar passive portfolio of indices. Alternatively, it can be assumed that a mutual fund is exposed to one investment style only that can be estimated from the data (see, Brown and Goetzmann [1997]).

In this chapter we focus on the performance of Dutch mutual funds over the period January 1990 through June 1997. We take equity as well as fixed income funds into account, and we will evaluate the performances of these funds on a relative as well as on a risk-adjusted basis. We will consider the question whether a Dutch investor can extend his mean-variance efficient set by investing in Dutch mutual funds with respect to a set of passive indices. Moreover, we show under which assumptions relative performance evaluation and performance evaluation on a risk-adjusted basis lead to similar conclusions about the potential ability of the fund managers.

The remainder of this chapter is organized as follows. In Section 4.2 we motivate the use of style analysis in mutual fund performance evaluation. Section 4.3 analyzes the investment styles and the relative performance of the sample of Dutch mutual funds. Moreover, we present some descriptive statistics for the sample of Dutch mutual funds that we employ. Section 4.4 evaluates fund performances on a risk-adjusted basis and answers the question whether mean-variance investors can improve the risk-return trade-off by taking a position in a Dutch mutual fund. In Section 4.5 we take short sales constraints into account, and we analyze the impact of these constraints on mutual fund performance evaluation. Finally, Section 4.6 concludes.

4.2 Return-based Style Analysis

Portfolio managers are often restricted to hold assets in a well-defined number of asset classes and are frequently limited to little or no leverage. One of the key determinants of a mutual funds' return is the asset allocation of the manager. For instance, for a mutual fund that primarily invests in equity this can mean that the management has to decide about the sectorial and regional allocation of the stocks and on the part to invest in growth stocks and the part to invest in value stocks.

As stressed by Brown and Goetzmann [1997], the self-reported investment style of the mutual fund does not always correspond to the actual investment behaviour. Consequently, in relative performance evaluation some mutual fund returns are compared with benchmark asset returns that do not correspond to the fund's actual investment style, possibly leading to a better relative mutual fund performance. As shown by, for instance, Sirri and Tufano [1997], individual investors select funds on prior performance information, investing more in funds that performed well over the last period. It is hard to judge whether fund managers are gaming relative performance evaluation on purpose or that there are other reasons for the observed misclassification. However, the impact on performance evaluation is the same.

Return-based style analysis (see, e.g. Sharpe [1992]), is an instrument to determine the exposure of a mutual fund to a number of major asset classes. To accomplish this task the following asset class model can be used

$$r_{t+1} = a + \sum_{k=1}^{K} b_k R_{kt+1} + u_{t+1}, \tag{4.1}$$

where r_{t+1} denotes the return on a mutual fund, K is the number of asset class factors, b_k is the sensitivity of r_{t+1} to the factor-mimicking portfolio R_{kt+1} and u_{t+1} is the idiosyncratic fund return, independent of all factor-mimicking returns. One of the main characteristics of asset class models is that the sensitivities are required to sum to 1, and should be larger than or equal to zero. The first characteristic implies that $\sum_{k=1}^{K} b_k R_{kt+1}$ can be interpreted as the return on a passive portfolio with the same investment style as the mutual fund. The second characteristic reflects the short selling restrictions often present for mutual fund managers. The constant a can be interpreted as the average tracking error between the mutual fund and the passive portfolio.

The primary goal of style analysis is not to evaluate a mutual fund's performance but to find a mimicking strategy that corresponds to the investment style of the mutual fund as closely as possible. After having determined this strategy, the mutual fund return in the subsequent periods can be compared with this passive strategy. In that way, a part of the fund's return can be assigned to investment style and a part can be assigned to active selection of the management. Therefore, style analysis can be used to circumvent the problem of self-reported styles in relative performance evaluation, and moreover, can be accomplished using return data only.

One of the aims of performance evaluation is to detect whether the fund manager has certain abilities that makes the fund an attractive investment product, such that an investor can extend his efficient investment set by taking a position in the mutual funds under consideration. Suppose that the return of a mutual fund can be written as

$$r_{t+1} = w' R_{t+1} + \alpha_{t+1}, \tag{4.2}$$

where $\alpha_{t+1} \sim N(\alpha, \Sigma_\alpha)$ reflects the ability of the fund manager, R_{t+1} is a K-dimensional vector of returns of asset classes and w is the corresponding weight vector.

Reconsider the following regression equation:

$$r_{t+1} = a + BR_{t+1} + u_{t+1}, \tag{4.3}$$

where B is a row vector of exposure coefficients to the K initial assets or asset classes and u_{t+1} is the idiosyncratic error term that is uncorrelated with all K asset class returns. The constant a is the parameter of interest and serves the purpose of measuring the potential ability of the fund manager. If we assume that the ability of the fund manager, α_{t+1}, is independent of the return on the K asset classes R_{t+1} then it is straightforward to show that B in (4.3) can be written as

$$B = \frac{Cov[w'R_{t+1} + \alpha_{t+1}, R_{t+1}]}{Var[R_{t+1}]} = w' \tag{4.4}$$

implying that B reflects the weights in the asset classes. Moreover, under the same assumption,

$$a = E[r_{t+1}] - BE[R_{t+1}] = E[\alpha_{t+1}] = \alpha \tag{4.5}$$

and that is the ability we are interested in. Consequently, under the assumption that the ability of the fund manager is independent of the return on the asset classes, return-based style analysis is an appropriate way to identify the potential ability of the fund manager.

4.3 Relative Performance Evaluation

The database that we employ contains 289 Dutch mutual funds, equity as well as fixed income and other types of funds, and is provided by Micropal Inc. Following previous studies on performance evaluation, we concentrate primarily on equity and fixed income funds. The sample that we analyze starts in January 1990. In Table 4.1 we present the main investment regions for the sample of funds as well as the number of funds with self-reported investment style corresponding to these investment regions. It appears that since the end of the eighties the number

Table 4.1: **Number of Funds per Investment Category and Size in guilders.** The table reports the number of mutual funds in existence before year t per investment category, and the the amount to manage, as reported in December 1996.

year investment region	1990	1993	1997/06	Size (in billion guilders)
European Equity	5	9	10	2.9
Regional/Country	9	13	33	11.8
North American Equity	7	9	11	1.0
Netherlands Equity	6	10	43	8.7
International Equity	14	18	31	22.5
European Bonds	4	9	18	2.8
Netherlands mix/balanced	5	9	11	2.4
Netherlands Guilder Bonds	8	21	45	12.5
International mix/balanced	10	13	15	5.2
International Bonds	13	17	24	23.2
total	81	128	241	93

of funds has grown enormously. During the last four years e.g., the number of funds has almost doubled. The funds that primarily invest in the Netherlands is the largest group, i.e. 99 out of the total number of 241. At the end of 1996, the total amount to manage is 47 billion guilders for equity funds and 46 billion guilders for fixed income funds. The largest equity fund is Robeco (International Equity) and the largest fixed income fund is Rorento (International Bonds) with 10.5 and 7.4 billion guilders respectively under management at the end of 1996.

In Panel A of Table 4.2 we present the average monthly return per investment category for the sample of mutual funds. During the period 1990-1997/06 as well as the subperiod 1993-

Table 4.2: **Summary Statistics.** Panel A of the table presents the average monthly return as well as the minimum and maximum return per investment category. Panel B reports the average monthly returns and the corresponding standard deviations (between parenthesis) for the asset classes.

Panel A: average monthly returns per investment category						
period	1990-1997/6			1993-1997/6		
investment region	mean	min	max	mean	min	max
European Equity	0.76	0.20	1.10	1.77	1.40	2.78
Regional/Country	0.48	-0.15	1.36	1.17	0.44	1.58
North American Equity	1.24	0.69	2.10	1.58	0.98	2.72
Netherlands Equity	1.46	1.15	1.68	1.94	0.69	2.46
International Equity	0.90	-0.11	1.40	1.56	0.48	2.28
European Bonds	0.79	0.61	1.15	0.73	0.33	1.68
Netherlands mix/balanced	1.00	0.84	1.12	1.25	0.43	1.57
Netherlands Guilder Bonds	0.52	0.25	0.75	0.51	0.18	0.79
International mix/balanced	0.64	0.33	1.04	0.97	0.56	1.43
International Bonds	0.61	0.40	0.79	0.62	0.27	1.38
Panel B: average monthly returns and standard deviations asset classes						
MSCI europe (Seurope)	1.09	(4.23)		1.73	(3.55)	
MSCI world (Sworld)	0.89	(4.57)		1.58	(3.91)	
Sal. Brothers G7 bond (Bworld)	0.80	(3.29)		0.84	(3.18)	
CBS stock index (Sneth)	1.59	(3.91)		2.41	(3.78)	
CBS bond index (Bneth)	0.74	(0.98)		0.71	(1.04)	
3-month deposit (depos)	0.52	(0.20)		0.38	(0.12)	

1997/07, the funds that primarily invest in Dutch equity have the highest average monthly return, i.e. 1.46% and 1.94% respectively (approximately 17.5% and 23.3% annually).

In relative performance evaluation, the average return of a mutual fund is compared with a benchmark that corresponds to the investment style. Panel B of Table 4.2 reports the average monthly returns and the corresponding standard deviations for the following six asset classes over two sample periods: three equity style indices: MSCI World equities, MSCI Europe equities, CBS Netherlands equities, two bond style indices: Salomon Brothers G7 bond index, the CBS general bond index and a three-month Dutch currency deposit. Looking at the results in Table 4.2 it appears that on average, for instance, Dutch mutual funds with main investment objective 'Netherlands Equity' underperform the corresponding Dutch stock index by 0.13% over the period 1990-1997/06, while over the subperiod 1993-1997/06 this underperformance increases to 0.47%. However, the average fund with investment objective 'International Equity' hardly shows under- or outperformance of the MSCI World index.

An explanation for the underperformance of the mutual funds might be that the mutual funds hold a cash position as well. One of the reasons for having such a liquidity position is that mutual funds can quickly respond to investors who sell their share in the mutual fund, without having to sell a corresponding part of the mutual fund's portfolio immediately. Therefore, a direct comparison between the return on the fund and the return on the index corresponding to the self-reported style is somewhat unfair. Style analysis will be used to circumvent this problem. Moreover, as mentioned by Brown and Goetzmann [1997], fund managers can easily mislead relative performance evaluation by investing in assets classes or parts of the world that are not in accordance with their reported investment style.

In order to examine whether the reported investment style coincides with the actual style, we apply return-based style analysis using the six style indices introduced above. Note that these indices can be interpreted as factor-mimicking portfolios for the factors that drive asset returns. In Table 4.3 we present a weighted average exposure for the main investment categories which is obtained by estimating the constrained regression equation (4.1) for the sample of Dutch equity funds. The weight of a fund in the computation of the weighted average is determined by the size of the fund as reported at the end of 1996.

Table 4.3: **Estimated Exposures.** The table shows the weighted-average estimated style of Dutch equity funds (in the colums labelled 'avg(wgt)') for five self-reported investment categories over the periods 1990-1997/06 and 1993-1997/06. The weight of the fund is determined by the size of the fund at the end of 1996. The colums labelled 'max' report the maximum estimated exposure of a fund in this investment category to a style index, while the column 'nr' reports the number of funds in the investment category that are exposed to a style index. The estimate for a is in % per month, while the number of funds behind the estimate for a is the total number of funds in the investment category.

Investment Region	European equity			Regional equity			North American equity		
period 1990 - 1997/06									
	avg(wgt)	max	nr	avg(wgt)	max	nr	avg(wgt)	max	nr
a	-0.23	-0.06	5	0.03	0.17	9	0.40	1.19	7
Seurope	0.59	0.76	4	0.26	0.41	3	0.04	0.22	2
Sworld	0.02	0.03	2	0.51	1.00	9	0.32	0.59	5
Bworld	0.00	0.09	1	0.08	0.18	2	0.46	0.78	6
Sneth	0.27	0.57	5	0.13	0.42	3	0.14	0.18	4
Bneth	0.07	0.43	3	0.00	0.00	0	0.00	0.00	0
depos	0.05	0.23	5	0.02	0.17	3	0.05	0.94	2
R^2	0.69	0.82		0.54	0.63		0.43	0.68	
period 1993 - 1997/06									
a	0.05	1.17	9	-0.45	-0.06	13	0.10	2.35	9
Seurope	0.66	0.85	9	0.17	0.88	4	0.09	0.30	5
Sworld	0.09	0.84	7	0.77	1.00	10	0.20	0.91	7
Bworld	0.00	0.09	2	0.00	0.30	1	0.34	0.58	8
Sneth	0.15	0.32	5	0.06	0.31	3	0.30	0.44	7
Bneth	0.00	0.00	0	0.00	0.00	0	0.00	0.00	0
depos	0.10	0.54	8	0.00	0.13	4	0.08	1.00	5
R^2	0.75	0.94		0.39	0.63		0.58	0.77	

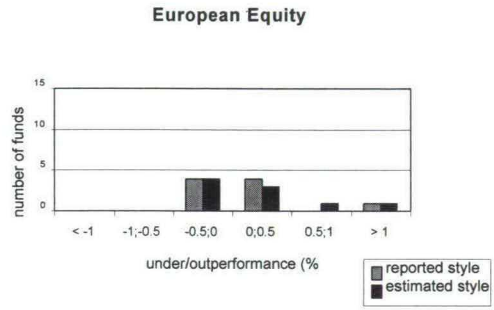
Table 4.3 continued

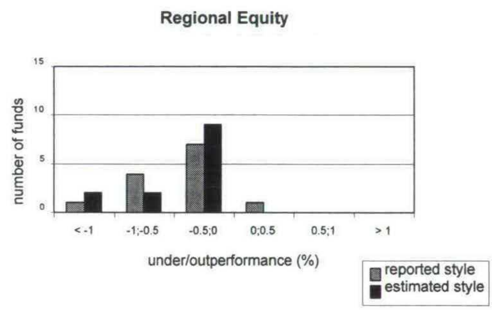
Investment Region	Netherlands equity			International equity		
period 1990 - 1997/06						
	avg(wgt)	max	nr	avg(wgt)	max	nr
a	0.22	0.29	6	-0.09	0.18	14
Seurope	0.12	0.17	6	0.09	0.36	13
Sworld	0.00	0.01	1	0.51	0.71	12
Bworld	0.00	0.03	1	0.01	0.40	6
Sneth	0.76	0.92	6	0.30	0.59	13
Bneth	0.03	0.08	2	0.01	0.23	5
depos	0.08	0.25	5	0.08	0.75	9
R^2	0.85	0.89		0.86	0.90	
period 1993 - 1997/06						
a	0.46	0.97	10	-0.10	0.56	18
Seurope	0.23	0.39	8	0.16	1.00	16
Sworld	0.05	0.13	7	0.53	0.76	16
Bworld	0.00	0.03	1	0.01	0.21	4
Sneth	0.50	0.74	10	0.22	0.50	14
Bneth	0.02	0.20	1	0.00	0.11	1
depos	0.20	0.73	9	0.09	0.73	15
R^2	0.72	0.92		0.82	0.88	

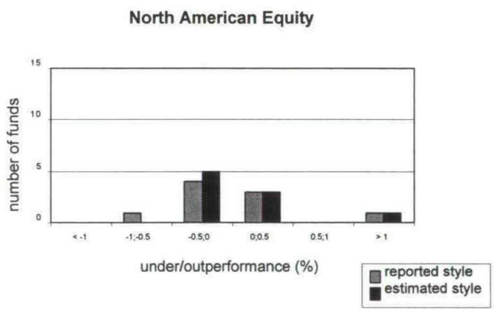
Not surprisingly, on average, the estimated maximum exposure within an investment category corresponds to the investment objective as reported by the mutual fund. However, if we look at funds with investment objective 'International Equity' then it appears that these funds are highly exposed to Dutch equity. Consequently, relative performance evaluation could in fact be gamed by the fund managers, since most of the mutual funds are overweighting Dutch equity. Furthermore, an individual investor that is considering to extend his portfolio with an internationally investing mutual fund, actually invests in a fund that is highly exposed to the domestic market. Moreover, if we compare the exposure of funds with investment objective 'Netherlands equity' over the two sample periods, then it appears that during the subperiod 1993-1997/06 mutual funds are more exposed to the cash deposit, indicating that mutual funds held more cash during this period. Consequently, the average monthly underperformance of 0.47% (Table 4.2) might be due to this rather large position in cash. The relatively low value of R^2 for the funds with objective 'Regional equity' and 'North American equity' indicates that the style indices that we employ do not sufficiently cover the actual investment style for these mutual funds.

In order to illustrate the effect of using style analysis in relative performance evaluation, we show in Figure 4.1 histograms for the relative performance of mutual funds over the period 1993 through 1997/06. In Figure 4.1, we compare relative performance evaluation using a benchmark that corresponds to the reported investment style with relative performance evaluation using the fund's estimated exposure to a passive portfolio of indices as benchmark. Note that our analysis differs from Brown and Goetzmann [1997], who assume that all funds invest

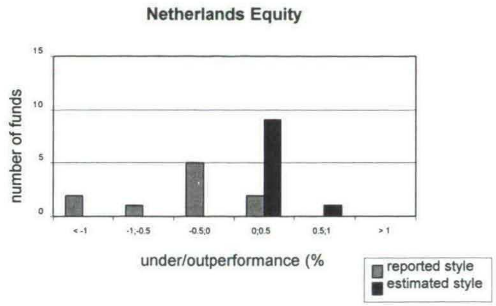
Figure 4.1: Relative Performance (%per month). The figures show the relative under or outperformance of the sample of mutual funds compared to the reported style or to the fund's estimated exposure to a passive portfolio of indices as benchmark. The evaluation period is 1993 - 1997/06.

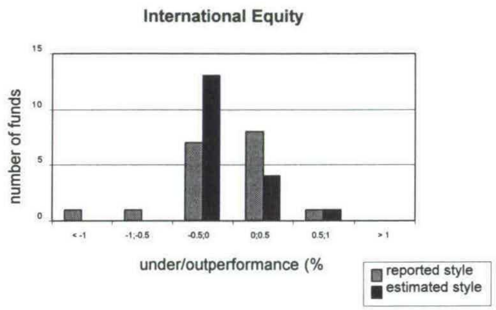






in one investment style only, that is estimated from the data. In general, it appears to be the case that the relative performance of mutual funds improves when using the estimated style of a fund as a benchmark. In particular, all the mutual funds with investment objective 'Netherlands equity' outperform the estimated style, while most of them underperform the Dutch CBS stock index, indicating that the cash position that the mutual funds hold seriously affects rela-





tive performance evaluation. As an illustration, at the individual fund level, one of the largest funds at the end of 1996 with objective 'Netherlands equity', i.e. the ABN AMRO Netherlands fund, outperformed the Dutch CBS stock index with only 0.04% per month, while after taking the exposure to different asset classes into account, the outperformance increases to almost 0.40% per month. The fund has an average estimated exposure to cash of almost 8% over the period 1993-1997/06.

For funds with investment objective 'International equity' the result is somewhat different. Although the extreme underperformance of a number of funds that was present in case of performance evaluation relative to the reported style has disappeared, most of the funds slightly underperform the replicating strategy that covers the fund's investment style, while the largest part of this group outperformed the benchmark corresponding to the reported investment style, i.e. thirteen vs nine funds respectively. For these funds some gaming might be present, probably caused by the exposure to Dutch equity.

For Dutch fixed income mutual funds we repeated the above style analysis. The result can be found in Appendix 4.A. Similar to the equity funds, it appears that most fixed income mutual funds have an estimated style that corresponds to the self-reported investment style.

However, the funds with objective 'International bonds' are heavily exposed to the Dutch CBS bond index. Consequently, a Dutch investor that considers to extend his initial portfolio with an internationally investing fixed income fund, is actually looking at funds that are highly exposed to the Dutch fixed income market.

4.4 Performance Evaluation using the Jensen Measure

One of the best known traditional ways of measuring mutual fund performances is the Jensen measure (Jensen [1968]). The generalized Jensen measure can be obtained as the intercept in the following regression equation:

$$(r_{t+1} - \eta) = \alpha_J(\eta) + B(R_{t+1} - \eta\iota_K) + v_{t+1}, \tag{4.6}$$

where r_{t+1} is the return vector of the mutual fund, R_{t+1} is a K-dimensional return vector of some benchmark assets, ι_K is a K-dimensional vector of ones, η is the risk free rate (if available) or some prespecified zero beta rate[1] associated with the investor's portfolio with expected return μ, B is a $1 \times K$ row vector of slope coefficients and v_{t+1} is the idiosyncratic error term which is uncorrelated with the K benchmark assets and has expectation zero. Note that $\alpha_J(\eta)$ generalizes the original alpha-measure proposed by Jensen [1968]. First of all, Jensen assumed that one of the benchmark assets is a risk free deposit, which implies that the zero beta rate equals the risk free rate. Secondly, Jensen considered the case of only two benchmark assets, i.e. the risk free rate and the market portfolio, while K in (4.6) is not restricted to two.

In a test for outperformance of a mutual fund with respect to K benchmark assets, the following hypothesis is tested:

$$H_0 : \alpha_J(\eta) = 0. \tag{4.7}$$

Alternatively, testing the hypothesis (4.7) can be interpreted as testing whether an investor with a particular risk aversion cannot extend the mean-variance efficient set by investing in the mutual fund, where η is the zero beta rate of the investor's portfolio corresponding to the investor's risk aversion. Or put differently, testing of (4.7) can be interpreted as testing for intersection of the initial and extended mean-variance frontier in the investor's initial optimal portfolio location. As shown by, for instance Elton, Gruber and Blake [1996], a positive Jensen measure indicates that a mean-variance investor whose current portfolio is covered by the K

[1] The zero beta rate η is the return on a portfolio that is uncorrelated with the investor's initial portfolio, and is therefore related to the investor's risk aversion.

benchmark assets can extend the mean-variance efficient set by taking a long position in the fund under consideration. A negative measure implies an extension by taking a short position in the fund.

It is important to note that there is a close link between the generalized Jensen measure and performance evaluation relative to the investment style estimated from return-based style analysis as in Section 4.3. It is straightforward to show that the generalized Jensen measure can be written as

$$\alpha_J(\eta) = a - (1 - B\iota_K)\eta, \tag{4.8}$$

where a and B can be obtained from the Huberman and Kandel [1987] regression equation (4.3). Recall that in return-based style analysis, equation (4.3) is estimated under the assumption that $B\iota_K = 1$. This implies that in return-based style analysis we already impose a part of the null hypothesis (4.7). Moreover, the restriction that the individual exposure coefficients should be equal or greater than zero is assumed to hold in the population.

Let us consider the case that where (4.2) holds, i.e. the fund manager's ability is uncorrelated with all style indices. In that case the generalized Jensen measure reduces to the parameter a which serves the purpose of measuring the ability of the fund manager as shown in (4.5). Consequently, the generalized Jensen measure has the advantage that it is directly related to efficient portfolio choice, but reduces to relative performance evaluation in an important special case. Note that in evaluating performances using the generalized Jensen measure it is usually not assumed that the ability of the fund manager is uncorrelated with the return on the K benchmark assets, and in that sense, the generalized Jensen measure is less restrictive in detecting a manager's ability.

In Table 4.4 we report the weighted-average generalized Jensen measure for the five investment categories, where we assume that the investor's initial portfolio is an efficient combination of the six asset class indices of Section 4.3. The weight of a fund in the computation of the weighted-average is, as before, determined by the size of the fund as reported at the end of 1996. We do not impose that an investor is obligated to invest all his wealth, implying the availability of a risk free asset with zero return[2]. For the sample period 1990-1997/06, we consider two different expected returns on the investor's portfolio. First of all, we consider an expected return of 0.55% monthly on his current efficient portfolio of K benchmark assets, which corresponds with a zero beta rate of 0.00%. This portfolio is the zero tangency portfolio, i.e. for expected returns greater than or equal to 0.55% all wealth is invested in the risky assets.

[2] It is not allowed to take a short position in a risk free asset with zero return.

Second, we consider an expected return of 3.5% on the investor's portfolio, corresponding to a zero beta rate of 0.534%. This zero beta rate corresponds to the intercept of the asymptote of the mean-variance frontier of the benchmark assets. Alternatively, it can be stated that this zero beta rate reflects the behavior of a risk neutral investor. We also computed the generalized Jensen measure for the subperiod 1993-1997/06. For this case, the zero tangency portfolio has an expected return of 0.40%, while an expected return of 3.5% on the investor's efficient portfolio now corresponds to a zero beta rate of 0.38%.

Table 4.4: **Generalized Jensen measure.** The table reports for two sample periods and a number of different expected returns on the investor's portfolio a weighted-average generalized Jensen measure with corresponding standard deviation (in the columns labelled 'avg(wgt)') as well as the minimum and maximum generalized Jensen measure per investment category. The weight of the fund in the weighted-average is determined by the size of the fund at the end of 1996. The standard deviation of the weighted-average is calculated by taking the correlations between the individual funds into account. All the values are on a monthly basis. A Jensen measure printed in *italics indicates significant at the 5% level.*

Investment	$\mu = 0.55\%$				$\mu = 3.50\%$			
Region	avg(wgt)		min	max	avg(wgt)		min	max
	α_J		α_J	α_J	α_J		α_J	α_J
period 1990 - 1997/06								
Europe	0.40	(0.64)	-0.14	0.65	-0.28	(0.22)	-1.51	-0.06
Regional	*-2.67*	(1.14)	*-3.75*	0.20	-0.16	(0.40)	*-1.27*	0.13
North America	0.39	(0.81)	-1.14	*9.57*	0.40	(0.28)	-0.23	1.46
Netherlands	0.58	(0.41)	-0.27	*1.23*	0.21	(0.14)	-0.15	0.29
International	-0.16	(0.39)	-0.91	1.11	-0.08	(0.14)	-0.91	0.18
period 1993 - 1997/06								
	$\mu = 0.40\%$				$\mu = 3.50\%$			
Europe	0.43	(0.56)	-4.10	*1.73*	0.07	(0.19)	*-0.28*	0.90
Regional	*-4.60*	(2.25)	*-5.81*	-1.39	-0.54	(0.75)	*-1.53*	0.13
North America	*2.53*	(0.81)	0.56	*12.66*	0.14	(0.27)	-0.38	*3.60*
Netherlands	*1.05*	(0.51)	-0.24	2.98	*0.49*	(0.17)	0.10	*1.06*
International	0.45	(0.59)	-7.88	*1.60*	-0.11	(0.20)	-0.45	0.52

From the estimates based on the period 1990-1997/06 it appears that, on average, the mutual funds with investment objective 'Netherlands equity' offer the investor an extension of the efficient set by taking a long position in most of the mutual funds under consideration, independent of the investor's risk aversion. These mutual funds show outperformance of the set of K benchmark assets, while the funds with investment objective 'Regional' or 'International equity' mostly underperform the K benchmark assets. For the estimates based on the subperiod 1993-1997/06 a comparable pattern is found. Independent of the investor's risk aversion, most of the mutual funds with investment objective 'Netherlands' offer the investor an extension of the efficient set by taking a long position in the funds under consideration.

The observed outperformance on a relative as well as on a risk-adjusted basis of Dutch mutual funds that mainly invest in the Netherlands might be an indication of a domestic market effect. The knowledge of the stock market where the fund is located might explain why this particular group of fund managers appears to have ability. Note that underperformance corresponds with taking a short position in the mutual funds for an extension of the efficient set. However, an investor will be confronted with short sales restrictions. Moreover, it is straightforward to show that a positive generalized Jensen measure for a mutual fund can coincide with short positions in the benchmark assets, which is an important drawback of this measure of outperformance.

At the individual fund level the result of the EDCC Netherlands Antilles fund (North American equity) is most striking. For the zero tangency portfolio the fund shows a significant extreme outperformance of about 9.5% and 12.5% on a monthly basis, for the period 1990-1997/06 and the subperiod 1993-1997/06 respectively. The outperformance of this fund also explains the average outperformance of the group of funds with investment objective 'North American equity'. When we take into account the estimated exposure to the different investment style indices that we obtained in Section 4.3 for this fund, it appears that the fund is highly exposed to the three-month currency deposit. Since the currency deposit is included in the set of K benchmark assets, the extreme outperformance cannot be explained by the huge exposure to this three-month currency deposit. Apparently, it is the case that the EDCC Netherlands Antilles fund is not a traditional mutual fund that is limited to little or no leverage (see also Fung and Hsieh [1997]). Consequently, the risk involved for a position in a so-called hedge fund is on average much higher than for a traditional fund. For the EDCC Netherlands Antilles fund the risk, as measured by the standard deviation, appears to be almost 10%, which is twice the risk involved in a position in the other funds within the same investment style.

4.5 Performance Analysis under Short Sales Restrictions

Since it is usually not possible for an individual investor to take a short position in a mutual fund or a benchmark asset, it is relevant to consider tests for outperformance of mutual funds under short sales constraints on the mutual funds as well as on the benchmark assets, with the exception of the three-month currency deposit. Markowitz [1991] has shown that the mean-variance frontier of all benchmark assets and the mutual fund under consideration consists of P segments where different assets have binding short sales restrictions. As shown by DeRoon, Nijman and Werker [1998], a test whether the efficient set can be extended by also investing in the mutual fund for an investor with zero beta rate η and short sales constraints, can be implemented by estimating the following regression equation

$$(r_{t+1} - \eta) = \alpha_J^{(p)}(\eta) + B^{(p)}(R_{t+1}^{(p)} - \eta\iota_L^{(p)}) + v_{t+1}^{(p)}, \tag{4.9}$$

where the superscript (p) means that the regression is based on the subvector $R_{t+1}^{(p)}$ of R_{t+1} that corresponds to the assets that are actually in the investor's portfolio, i.e. for which the short sales constraints are not binding. A test for extension of the efficient set is equivalent to testing whether the following hypothesis:

$$\alpha_J^{(p)}(\eta) \leq 0, \tag{4.10}$$

holds. This one-sided inequality constraint (4.10) can be tested using the Wald test under inequality constraints (see, e.g. Kodde and Palm [1986]).

Similar to the unrestricted case, we test the hypothesis for a number of expected returns on the investor's efficient portfolio. For the sample period 1990 - 1997/06 the investor's initial efficient portfolio under short sales restrictions consists of three assets, i.e. CBS stock index, CBS bond index and a three-month deposit, independent of the investor's risk aversion. For the sample period 1993-1997/06, the initial portfolio of assets also consists of the Morgan Stanley Europe index for risk aversions corresponding to a zero beta rate of 0.00%. In Table 4.5 we report the weighted-average generalized Jensen measure under short sales restrictions for five investment categories for different expected returns. For the sample period 1990-1997/06 we test for outperformance for an expected return of 0.54% associated with the zero tangency portfolio, and for an expected return of 3.5%. Recall that the zero beta rate for this portfolio with a rather high expected return can be obtained as the intercept of the asymptote of the mean-variance frontier of the CBS stock index, CBS bond index and a three-month deposit.

If it is not possible for an individual investor to take a short position in a mutual fund, the negative numbers in the table indicate that the investor cannot extend the efficient set by taking a position in these funds. It appears that the mutual funds that primarily invest in 'Netherlands equity' show outperformance of the benchmark assets, even after imposing short sales restrictions. Usually, it is found that only a small number of funds have a positive Jensen measure in a frictionless market (see, e.g. Malkiel [1995], Gruber [1996]). Although the outperformance is only significant for one out of ten mutual funds in a market with short sales constraints, the fact that for most of these funds the generalized Jensen measures is positive, is remarkable and might be due to a domestic market effect.

In order to illustrate the effect of imposing short sales restrictions on the benchmark assets in more detail, we show in Figure 4.2 the generalized Jensen measure without (unrestricted) as well as with imposing short sales restrictions (restricted). In both cases, we consider an

Table 4.5: **Generalized Jensen measure under short sales constraints.** The table reports for two sample periods and a number of different expected returns on the investor's portfolio a weighted-average generalized Jensen measure and corresponding standard deviation (in the colums labelled 'avg(wgt)') under short sales constraints as well as the minimum and maximum generalized Jensen measure per investment category. The weight of the fund in the weighted-average is determined by the size of the fund at the end of 1996. The standard deviation of the weighted-average is calculated by taking the correlations between the individual funds into account. All the values are on a monthly basis. A Jensen measure printed in *italics indicates significant at the 5% level.*

Investment Region	$\mu = 0.54\%$				$\mu = 3.50\%$			
	avg(wgt) α_J		min α_J	max α_J	avg(wgt) α_J		min α_J	max α_J
period 1990 - 1997/06								
Europe	0.09	(0.72)	-0.71	0.78	-0.53	(0.25)	-1.10	-0.35
Regional	-3.40	(1.29)	-4.40	-0.25	-0.67	(0.44)	-2.02	-0.44
North America	-0.37	(1.05)	-2.23	*9.31*	0.02	(0.36)	-0.79	0.73
Netherlands	0.52	(0.41)	-0.23	*1.12*	0.16	(0.14)	-0.14	0.21
International	-0.63	(0.63)	-1.26	1.27	-0.47	(0.22)	-0.15	-0.12
period 1993 - 1997/06								
	$\mu = 0.40\%$				$\mu = 3.50\%$			
Europe	0.37	(0.56)	-4.63	*1.67*	-0.25	(0.28)	-0.65	0.67
Regional	-5.56	(2.78)	-6.50	-1.30	-1.28	(0.98)	-2.04	-0.42
North America	*2.41*	(0.99)	0.59	*12.41*	-0.28	(0.33)	-0.97	*3.10*
Netherlands	*1.01*	(0.50)	-0.31	*2.95*	*0.37*	(0.18)	0.00	*0.90*
International	0.13	(0.79)	-7.70	*1.37*	-0.44	(0.29)	-0.77	0.26

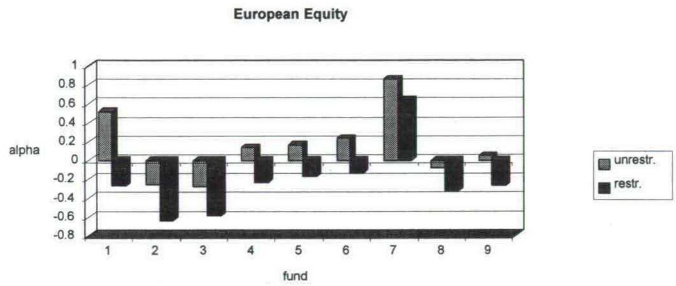
expected return of 1.5% per month on the investor's current efficient portfolio of benchmark assets. Note that in the unrestricted case the investor's initial portfolio consists of K benchmark assets, while in the restricted case the initial portfolio only consists of the CBS stock index, CBS bond index and a three-month currency deposit. Moreover, in the unrestricted case an expected return of 1.5% corresponds to a zero beta rate of 0.379%, while in the case of short sales constraints the zero beta rate equals 0.384%.

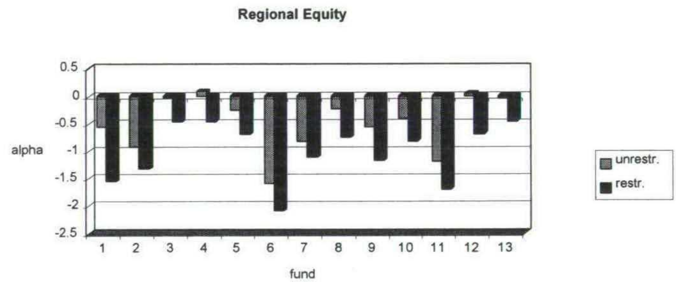
A striking result from Figure 4.2 is that the average performance of all fund categories decreases when short sales constraints on the benchmark are taken into account. This must be due to the fact that in the unrestricted case the investor could have extended the efficient set by short selling a number of the benchmark assets. Apparently it is the case that the short sales restrictions together with the correlation of the mutual funds with the remaining benchmark assets leads to a decrease in the generalized Jensen measure.

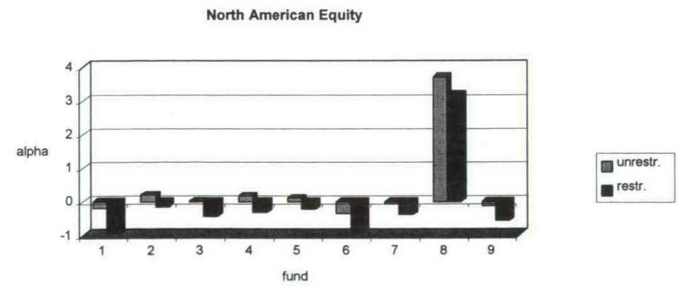
4.6 Concluding Remarks

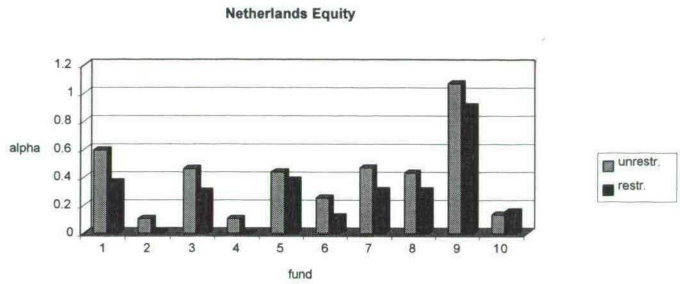
Returns-based style analysis is useful for improving relative performance evaluation of mutual funds. Since the actual investment style of a mutual fund does not necessarily correspond with the self-reported investment style, simply comparing realized fund returns with benchmarks

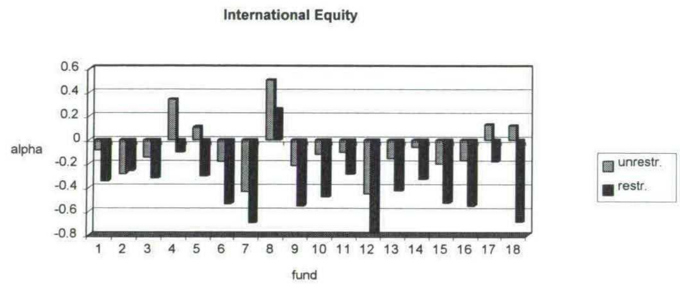
Figure 4.2: **Generalized Jensen measure (%per month).** The figures show the generalized Jensen measure without (unrestricted) as well as with imposing short sales restrictions (restricted). The evaluation period is 1993 - 1997/06.











corresponding to the reported styles might be misleading. In this chapter we have shown that style analysis can be used to objectively determine a fund's actual investment style. For most of the funds in the sample of Dutch mutual funds that we employ in this chapter, it appears that the maximum estimated exposure to an investment style index corresponds with the self-reported style. However, a large number of the funds are also highly exposed to other style indices, such as a three-month currency deposit, indicating that the funds hold a large cash position as well. In a few cases, we find that the fund is more exposed to cash than to the self-reported style. In these cases the mutual fund is probably not a real mutual fund that is restricted to little or no leverage.

In order to answer the question whether Dutch investors can extend the efficient set by taking a position in a Dutch mutual fund, we evaluated the performances of the fund on a relative basis as well as by using the generalized Jensen measure. If we take into account the exposure to style indices, the relative performance increases for most of the mutual funds. In particular, the funds that mainly invest in 'Netherlands equity' outperform the passive portfolio reflecting the fund's investment style, while underperformance dominates using the self-reported style as benchmark. It appears that most of the funds in the sample show underperformance, however, the group of funds with investment objective 'Netherlands equity' mainly shows outperformance. This finding is robust for incorporating short sales restrictions on most of the benchmark assets, and for the risk attitude of the investor. Consequently, most mean-variance investors can extend the efficient set by taking a long position in Dutch mutual funds that mainly invest in 'Netherlands equity'.

A final point to note is that relative performance evaluation is more restrictive in detecting potential ability of a fund manager than risk-adjusted performance evaluation. Under the assumption that the manager's ability is independent of the return on the benchmark assets, relative performance evaluation is an appropriate method to evaluate mutual funds. Moreover, if the exposure to different asset classes is not known, style analysis can be used as the instrument to determine these exposures.

Appendix 4.A

Style Analysis: Fixed Income Funds

Table 4.6: **Estimated Exposure Fixed Income Funds.** The table reports the weighted-average estimated style of Dutch fixed income funds (in the colums labelled 'avg(wgt)') for five self-reported categories over the periods 1990-1997/06 and 1993-1997/06. The weight of the fund in the weighted-average is determined by the size of the fund at the end of 1996. The colums labelled 'max' report the maximum estimated exposure of a fund to a style index, while the column 'nr' reports the number of funds that are exposed to a style index. The estimate for α is in % per month, while the number of funds behind the estimate for α is the total number of funds in the investment category.

Investment Region	European bonds			Netherlands mix			Netherlands bonds		
period 1990 - 1997/06									
	avg(wgt)	max	nr	avg(wgt)	max	nr	avg(wgt)	max	nr
α	-0.05	-0.03	4	0.02	0.07	5	-0.12	0.02	8
Seurope	0.02	0.10	2	0.01	0.03	2	0.00	0.02	1
Sworld	0.03	0.38	3	0.01	0.03	2	0.01	0.04	3
Bworld	0.05	0.12	4	0.08	0.28	5	0.02	0.08	4
Sneth	0.03	0.41	2	0.38	0.41	5	0.01	0.09	4
Bneth	0.84	0.94	4	0.32	0.40	5	0.80	0.95	8
depos	0.03	0.49	3	0.20	0.26	5	0.16	0.54	8
R^2	0.66	0.92		0.77	0.93		0.73	0.86	
period 1993 - 1997/06									
α	-0.17	0.05	9	0.12	0.44	9	-0.13	0.09	21
Seurope	0.06	0.11	6	0.08	0.35	7	0.00	0.12	2
Sworld	0.02	0.44	4	0.08	0.15	5	0.00	0.01	1
Bworld	0.07	0.14	6	0.07	0.15	8	0.02	0.18	17
Sneth	0.02	0.44	4	0.30	0.38	9	0.02	0.06	13
Bneth	0.53	0.91	9	0.15	0.41	7	0.70	0.95	20
depos	0.30	0.51	7	0.33	0.57	9	0.25	1.00	19
R^2	0.59	0.92		0.70	0.93		0.75	0.86	

Table 4.6 continued

Investment Region	International mix			International bonds		
period 1990 - 1997/06						
	avg(wgt)	max	nr	avg(wgt)	max	nr
a	-0.18	0.02	10	-0.10	0.01	13
Seurope	0.03	0.19	9	0.01	0.11	5
Sworld	0.18	0.39	10	0.06	0.12	7
Bworld	0.02	0.10	7	0.10	0.89	13
Sneth	0.14	0.33	9	0.01	0.08	9
Bneth	0.32	0.42	10	0.67	0.86	11
depos	0.32	0.70	10	0.15	0.65	11
R^2	0.70	0.91		0.64	0.82	
period 1993 - 1997/06						
a	-0.14	0.14	13	-0.20	-0.05	17
Seurope	0.09	0.35	11	0.03	0.20	12
Sworld	0.09	0.44	13	0.04	0.97	6
Bworld	0.05	0.13	10	0.12	0.83	17
Sneth	0.17	0.25	11	0.01	0.15	10
Bneth	0.33	0.49	10	0.52	0.88	13
depos	0.27	0.65	12	0.27	0.72	13
R^2	0.80	0.90		0.63	0.83	

Chapter 5

Estimating Short-Run Persistence in Mutual Fund Performance

This chapter analyzes the properties of a number of estimators that can be used to estimate short-run persistence in mutual fund returns. When data for different funds are pooled, it is advisable to correct for cross-sectional differences in expected returns. However, these adjustments may induce biases in the estimated persistence coefficients and thus lead to spurious persistence. Theoretical derivations, combined with a Monte Carlo study, show that the importance of these biases cannot be neglected for the samples that are typically used in applied work, in particular if the number of time periods is small. We also estimate the short-run persistence in two samples of U.S. open-end mutual funds using quarterly returns for 1986-1994. An important conclusion is that the results are quite sensitive to the estimation method that is employed.

5.1 Introduction

The fast growing mutual fund industry tries to attract investors by advertising its past record of fund returns. Empirical evidence (see Patel, Zeckhauser and Hendricks [1992]) shows that investors are more willing to invest money in a mutual fund if the fund returns are high compared to other mutual funds. Apparently, these investors expect that mutual funds with above average returns in one period will continue to have above average returns. If this is indeed the case, an investment strategy based on identifying funds with so-called hot hands can increase the expected return on investors' portfolios of mutual funds (see, e.g. Hendricks, Patel and Zeckhauser [1993]).

One approach to estimate persistence in mutual fund returns is based on regressions of a sample of funds' current returns upon a range of lagged returns. To adjust for the fact that market equilibrium returns are potentially different for the different funds, attempts are usually made to eliminate this cross-sectional variation by subtracting some measure of "expected returns" from the left hand side regression variable. However, several approaches use lagged returns in the calculation, such that a mechanical relation between risk-adjusted returns and lagged returns arises. Consequently, these methods may induce spurious findings of short-run persistence.

In this chapter we analyze this problem in more detail. When expected returns are constructed as the mean return over the sample period or as an estimated factor model expected return (as in Hendricks, Patel and Zeckhauser [1993]), the induced biases are negative and decreasing in T, the number of time periods that is used to estimate the expected returns. This means that these biases cannot generate a spurious finding of hot hands in mutual funds. On the contrary, it will decrease empirical evidence in favor of hot hands. An approach suggested by Jegadeesh [1990], using future returns to estimate "expected returns", or a pooled instrumental variables approach, suggested below, avoid the problem. In this chapter, we analyze the performance of these alternative estimators using analytical derivations and a Monte Carlo study. Furthermore, we use two samples of quarterly returns on U.S. open-end mutual funds for the period 1986-1995 to analyze the estimated persistence of mutual fund performance as obtained by the alternative estimators. The subsample of growth funds appears to have a persistence pattern that is quite similar to the one found by Hendricks, Patel and Zeckhauser [1993] for the period 1974-1988.

The remainder of this chapter is organized as follows. In Section 5.2 we present five methods, mostly proposed in the literature, to estimate patterns of persistence in mutual fund returns. For a finite number of periods, several of these methods can be shown to have an asymptotic bias in the estimated coefficients. In Section 5.3 we derive analytical expressions for these biases, starting from the hypothesis that the returns on each fund are independent drawings from a time-invariant distribution. To simplify our expressions, we only consider the case where one lag is included in the persistence equation. For the general case, with a larger number of lags, we present additional results in Section 5.4 using a Monte Carlo study. The results from this study indicates that the analytical results from Section 5.3 are equally valid for other lags than the first one. In addition, we consider the case where fund returns do have a pattern of predictability and discuss to what extent the estimation methods are able to detect and estimate this pattern. Section 5.5 presents the results of an empirical study into the short-run persistence in a sample of open-end mutual funds, selected from the Morningstar database, over the period 1986 to 1994. Finally, Section 5.6 concludes.

5.2 Persistence of Returns

Active selection among mutual funds can be profitable if mutual fund returns show a pattern of predictable behavior. If this is the case, the expected return on an investor's mutual fund portfolio can be increased if he is able to identify funds that will be superior performers in the future. For instance, if funds exhibit significant positive persistence in returns over a certain period then it can be worthwhile for an investor to select the funds with a high return, relative to their own unconditional mean return, over that period to increase the expected return on his portfolio. However, before we can test an economically valuable investment strategy, we first have to identify the form of the pattern of predictable returns.

For the moment, let us consider N mutual funds with an observed return history of T periods. Furthermore, we assume that the conditional expected return of mutual fund i in period t can be written as

$$E_{t-1}[r_{it}] = \gamma_{i0} + \sum_{j=1}^{J} \gamma_{ij} r_{i,t-j} = \mu_i + \sum_{j=1}^{J} \gamma_{ij}(r_{i,t-j} - \mu_i). \tag{5.1}$$

where r_{it} is the return in excess of the risk free rate and $\mu_i = E[r_{it}] = \gamma_{i0}/(1 - \sum_j \gamma_{ij})$ is the unconditional expected excess return. The coefficients γ_{ij} $(j = 1, ..., J)$ reflect persistence in the excess return of fund i, relative to its own unconditional mean. Clearly, the efficient market hypothesis implies that each parameter γ_{ij} is equal to zero. Recent empirical evidence (see Grinblatt and Titman [1992], Hendricks, Patel and Zeckhauser [1993], Goetzmann and Ibbotson [1994], Carhart [1997a]) indicates that there may be some, statistically significant, short-run persistence in mutual fund returns. For example, Hendricks, Patel and Zeckhauser [1993] claim that the predictable behavior of mutual fund returns can be profitable for an investor who actively selects mutual funds according to certain investment strategies based upon funds' past returns.

Predictable behavior of mutual fund returns can be estimated using regression analysis, after rewriting (5.1) as

$$r_{it} = \gamma_{i0} + \sum_{j=1}^{J} \gamma_{ij} r_{i,t-j} + \varepsilon_{it} \tag{5.2}$$

where ε_{it} is the unexpected return of fund i in period t. In principle, (5.2) can be estimated for each of the N funds in the sample. However, usually one is not directly interested in the persistence pattern of an individual fund, but rather in examining whether a group of mutual funds has, on average, a pattern of predictable returns. Moreover, individual estimates are likely to be very inaccurate due to a small signal-to-noise ratio, particularly when the fund's history is short. Therefore, it is common to pool the returns of all funds and estimate a set of common persistence coefficients, or, when homogeneity of γ_{ij} is not imposed, estimate a set of – hopefully – average coefficients. In the sequel, we shall consider several approaches that are suggested for this purpose.

A first way to estimate short-run persistence of mutual fund returns follows Fama and Macbeth [1973] and is based on cross-sectional regressions of the form

$$r_{it} = k_t + \sum_{j=1}^{J} \gamma_{jt} r_{i,t-j} + u_{it}, \quad i = 1, ..., N, \tag{5.3}$$

where homogeneity of the persistence pattern over the funds is imposed, while variation over time is not excluded. In other words, it estimates the persistence of *relative* performance. This standard Fama-Macbeth procedure implies that (5.3) is estimated for each period t, after which parameter estimates, and standard errors, are obtained from the time series of regression estimates. In particular, the set of estimated slope coefficients is treated as a random sample from a population with constant mean γ_j. We shall refer to this approach as *FM*.

Essentially, (5.3) checks for autocorrelation in fund returns imposing that these are drawings from a distribution with a *common*, time-varying, mean. That is, the specification in (5.3) does not only impose that the predictability pattern is the same for all funds, but also that the expected return on each of the funds is same. As argued by Jegadeesh [1990], this may lead to biased estimates for the persistence coefficients, because, relative to the common mean, fund returns do exhibit correlation over time, even if all γ_{ij} are zero. Intuitively, funds with a high average return are simply more likely to have high returns (relative to the common mean) in all periods. Given that there is variation in expected returns over the funds, estimating (5.3) by ordinary least squares will find spurious correlations over time between current and past returns.

Most solutions for this problem try to eliminate γ_{i0} or (equivalently) μ_i by subtracting some estimate of it from the left hand side variable. Denoting this estimate by $M_{t-1}(r_{it})$, the resulting cross-sectional regression is given by

$$r_{it} - M_{t-1}(r_{it}) = k_t + \sum_{j=1}^{J} \gamma_{jt} r_{i,t-j} + \tilde{u}_{it}, \quad i = 1, ..., N, \tag{5.4}$$

which can be estimated according the Fama Macbeth procedure. A number of different estimators of the unconditional expectation have been proposed in the literature (see Jegadeesh [1990], Hendricks, Patel and Zeckhauser [1993]). Let us consider three possible choices for M_{t-1},

1. $M_{t-1}(r_{it}) := \bar{r}^*_{i,t} = \frac{1}{S}\sum_{s=1}^{S} r_{i,t+s}$, the average return over period $t+1$ to $t+S$ for some

positive S.

2. $M_{t-1}(r_{it}) := \bar{r}_i = \frac{1}{T}\sum_{s=1}^{T} r_{is}$, the average historical return over period 1 to T.
3. $M_{t-1}(r_{it}) := \hat{r}_{it} = b_{0i} + \sum_{k=1}^{K} \beta_i^k(\delta_{kt} - r_f)$, the return predicted by a linear K-factor model.

We shall refer to the estimation methods of equation (5.4) with the above specification of the unconditional expected return as *FM1*, *FM2* and *FM3*, respectively. While the first choice, corresponding to the one made by Jegadeesh [1990], indeed eliminates the bias due to variation in expected returns over the funds, the latter two, examples of the choices made by Hendricks, Patel and Zeckhauser [1993], generate another bias as we will show below, which may induce a spurious finding of negative short-run persistence in returns. A disadvantage of Jegadeesh' approach is that it requires returns over the period $t+1$ to $t+S$ to estimate the unconditional expected returns. In particular, when economically valuable investment strategies are investigated, this method does not seem very attractive because the number of time series observations available is often small.

As an alternative strategy, we suggest a different estimation method based on the analogy of removing fixed individual effects in a dynamic panel data model (see Hsiao [1985, p. 71 ff.], Baltagi [1995, p. 125 ff.]. In this approach, the returns over the N funds and T periods are pooled, after which the model is written in terms of first differences, while including a time-varying intercept. Although this eliminates the fund specific effects in γ_{i0}, it does lead to correlation between lagged returns and the error term, invalidating least squares estimation. Therefore, we follow Anderson and Hsiao [1981] and estimate the resulting equation

$$r_{it} - r_{it-1} = \gamma_t + \sum_{j=1}^{J} \tilde{\gamma}_j(r_{i,t-j} - r_{i,t-j-1}) + v_{it}, \quad t = 1, ..., T \text{ and } i = 1, ..., N \tag{5.5}$$

by instrumental variables. A valid instrument for $r_{i,t-1} - r_{i,t-2}$ is given by $r_{i,t-2}$, while all other regressors can be treated as exogenous and thus serve as their own instruments. We shall refer to this method as *pooled IV*.

The five estimation methods above all produce estimates for some average of the individual persistence coefficients over the funds, or for the common value of these coefficients when there is no fund heterogeneity in γ_{ij} ($j = 1, ..., J$). To show that some of these estimators may produce seriously biased estimates, we shall first, in the next section, derive analytical expressions for their probability limit when the number of lags in the regression is restricted to 1 ($J = 1$) and the true persistence coefficients are all equal to zero ($\gamma_{ij} = 0$ for all i). More general cases are considered in Section 5.4, on the basis of a Monte Carlo study.

5.3 Properties of the Estimators: Analytical Results

Deriving analytical expressions for the properties of the range of estimators discussed above in a very general case is tedious and does not provide much insight. Therefore, we simplify the analysis, and shall consider in this section the case where only one lag is included in the regressions ($J = 1$). For the moment, we shall also assume that the efficient market hypothesis holds, which implies that past returns do not have any indicative value for future returns. True fund returns are assumed to be generated by the following one factor model

$$r_{it} = \beta_i r_{mt} + \eta_{it} \tag{5.6}$$

where we shall refer to r_{mt} as the return on the market portfolio in excess of the risk free rate (although it may denote any other factor that prices the funds), and where β_i is the sensitivity of fund i with respect to the market portfolio. For a given fund i, the unobservable error terms η_{it} are assumed to be i.i.d. drawings from a distribution with zero mean and constant variance, independent of r_{mt}. Consequently, the data generating process (5.6) implies that the expected excess return on fund i, as defined in (5.1), is given by $\mu_i = \beta_i \mu_m$, where $\mu_m \neq 0$ is the expected excess return on the market portfolio. It also implies that $\gamma_{ij} = 0$ for all i and j. Although (5.6) may be somewhat restrictive, it serves our purpose as it implies that any excess performance is the result of luck (a good draw of η_{it}), and has no predictive power for future performance. The β_i's are assumed to be random drawings from a distribution with mean μ_β and variance σ_β^2, uncorrelated with η_{it} $(t = 1, ..., T)$.

Let us first consider the OLS estimators for γ_{1t} in (5.3) using the N fund returns in period t, which form the basis for the FM method. The pseudo true value[1] for the OLS estimator $\hat{\gamma}_{1t}$ is given by

$$\gamma_{1t}^* = \frac{\mathrm{Cov}_t[r_{it}, r_{i,t-1}]}{\mathrm{V}_t[r_{i,t-1}]}, \tag{5.7}$$

where the suffix t attached to the (co)variances is used to indicate *cross-sectional* (co)variances[2] for all funds that are available at time t. Note that in a cross-section at time t, the market returns in period t or before can be considered as given. Using the data generating process in (5.6), it can be shown that

$$\mathrm{Cov}_t[r_{it}, r_{i,t-1}] = \sigma_\beta^2 r_{mt} r_{m,t-1}, \tag{5.8}$$

[1] The pseudo true value of an estimator $\hat{\theta}_N$ is defined as the probability limit of that estimator when $N \to \infty$.
[2] Note that this is not the same as conditional (co)variances.

which can be either positive or negative, and that

$$\mathrm{V}_t[r_{i,t-1}] = \sigma_\beta^2 r_{mt}^2 + V_t[\eta_{it}]. \tag{5.9}$$

The result in (5.8) shows that the problem of cross-sectional correlation between r_{it} and $r_{i,t-1}$, even when r_{mt} and u_{it} are serially uncorrelated, is due to cross-sectional variation in expected returns over the funds. The *FM* estimate, obtained as the time-average of $\hat{\gamma}_{1t}$, also suffers from a non-zero pseudo true value (and thus a bias) as the average of (5.8) nor the average ratio of (5.8) and (5.9) is equal to zero. The bias in the *FM* estimator can be expected to be positive, as r_{mt}, though uncorrelated over time, will have a positive mean.

In order to eliminate the above bias, Jegadeesh [1990] suggests to adjust the left hand side of (5.3) by subtracting an unbiased estimate for the expected return[3], based on a moving average of S future returns. Alternatively, the sample average return, or the predicted value from the one-factor model can be used. This results in the methods referred to as *FM1*, *FM2* and *FM3*, respectively. The pseudo true value of the resulting estimators can be obtained by replacing the numerator in (5.7) by $Cov_t[r_{it} - M_{t-1}(r_{it}), r_{i,t-1}]$, with the appropriate choice for $M_{t-1}(r_{it})$. Ideally, $M_{t-1}(r_{it})$ is correlated with $r_{i,t-1}$ in such a way that the numerator in (5.7) equals zero (on average).

Let us now consider the pseudo true value of the OLS estimator for γ_{1t} in these three cases. Using the assumptions of the data generating process in (5.6), the following expression for the numerator can be derived for the *FM1* method

$$Cov_t[r_{it} - M_{t-1}(r_{it}), r_{i,t-1}] = \sigma_\beta^2 (r_{mt} - \bar{r}_{m,t}^*) r_{m,t-1} \tag{5.10}$$

where $\bar{r}_{m,t}^*$ denotes the average market return over the period $t+1$ to $t+S$ $(S > 0)$. Taking the expectation over t in this numerator gives zero, where we use the assumption that r_{mt} are independent drawings from a distribution with a constant mean μ_m and variance σ_m^2. Consequently, we do not expect a bias for this estimator.

However, in the *FM2* procedure, where the average return over the whole sample period is employed, we have

$$Cov_t[r_{it} - M_{t-1}(r_{it}), r_{i,t-1}] = \sigma_\beta^2 (r_{mt} - \bar{r}_m) r_{m,t-1} + \frac{1}{T} V_t[\eta_{it}] \tag{5.11}$$

which differs in two aspects from (5.10). First, the presence of an additional second term and second, the average market return $\bar{r}_m$ now also includes $r_{m,t-1}$. Consequently, taking the expectation over t in (5.11) gives a non-zero value. Furthermore, combining (5.11) with (5.9) and

[3] Due to a slightly different assumption on the data generating process, the expressions in Jagadeesh are similar but not identical to ours.

taking expectations over t in numerator and denominator, results in the following expression for the pseudo true value of the *FM2* estimator[4]

$$\gamma_1^* \approx -\frac{1}{T}\frac{\sigma_\beta^2\sigma_m^2 + V[\eta_{it}]}{\sigma_\beta^2(\sigma_m^2+\mu_m^2)+V[\eta_{it}]}, \tag{5.12}$$

which implies that the expected bias is negative and in absolute value somewhat less than $\frac{1}{T}$.[5]

Considering the data generating process (5.6), the *FM3* procedure is now based on the returns predicted by the linear one-factor model, i.e. $M_{t-1}(r_{it})$ is based on a time series regression of r_{it} on r_{mt}. Using the expression for the OLS estimators for the intercept term, one can write

$$M_{t-1}(r_{it}) = \bar{r}_i + \hat{\tau}_{1i}(r_{mt} - \bar{r}_m), \tag{5.13}$$

where

$$\hat{\tau}_{1i} = \frac{\sum_t (r_{it}-\bar{r}_i)(r_{mt}-\bar{r}_m)}{\sum_t (r_{mt}-\bar{r}_m)^2}. \tag{5.14}$$

From this, it follows that

$$\mathrm{Cov}_t[M_{t-1}(r_{it}), r_{i,t-1}] = \sigma_\beta^2 r_{mt} r_{m,t-1} + T^{-1}\mathrm{V}_t[\eta_{it}], \tag{5.15}$$

Combining this result with (5.8) gives

$$\mathrm{Cov}_t[r_{it} - M_{t-1}(r_{it}), r_{i,t-1}] = -T^{-1}V_t[\eta_{it}]. \tag{5.16}$$

Consequently, we can expect a slightly smaller bias (in absolute value) in the *FM3* estimator based on predicted returns from the factor model compared to the one based on average historical returns. The comparable expression to (5.12) is given by

$$\gamma_1^* \approx -\frac{1}{T}\frac{V[\eta_{it}]}{\sigma_\beta^2(\sigma_m^2+\mu_m^2)+V[\eta_{it}]}. \tag{5.17}$$

Finally, let us consider the *pooled IV* method. The pseudo true value of the IV estimator for γ_1 is now equal to

$$\gamma_1^* = \frac{\mathrm{Cov}[(r_{it}^* - r_{it-1}^*), r_{it-2}]}{\mathrm{Cov}[(r_{it-1}^* - r_{it-2}^*), r_{it-2}]} \tag{5.18}$$

where the covariances now denote covariance over all N funds and T time periods, and the starred returns denote returns in excess of the average return over all funds in the same period.[6]

---

[4] The approximation sign is due to the fact that we do not take expectations over the ratio but over numerator and denominator separately.

[5] Hendricks, Patel and Zeckhauser [1993] seem to encounter a bias of this magnitude in their bootstrap simulations discussed at the end of the paper (compare their footnote 22). They do not, however, adjust their claim that "the estimated slope coefficients are unbiased" (their Table 1).

[6] Transforming all variables like this is equivalent to including a time dummy for each period.

Considering the assumptions of the data generating process (5.6), the numerator in (5.18) equals

$$\mathrm{Cov}[(r^*_{it} - r^*_{it-1}), r_{it-2}] = \sigma^2_\beta \mathrm{E}[(r^*_{mt} - r^*_{mt-1})\, r_{mt-2}] = 0. \tag{5.19}$$

Thus, similar to the *FM1* method, we can expect a zero bias for this *pooled IV* estimator.

Recall that the five methods discussed above are used to estimate the predictable behavior in mutual fund returns. All methods give an estimate of the average persistence coefficient for the first lag. However, under our data generating process, any superior performance is due to luck, and is not an indication for future performance. Nevertheless, some of the methods discussed above do find a spurious pattern of persistence in returns. The size and sign of this asymptotic bias for the five estimation methods are summarized in Table 5.1. In case of the standard *FM*

Table 5.1: **Asymptotic bias in γ_1-estimates.** The table shows the expected sign and size of the asymptotic bias in the estimated first persistence coefficients, where T is the number of time series observations available per fund.

Estimation Method	Expected Size and Sign
FM	Bias > 0
FM1	Bias $= 0$
FM2	Bias $\approx -\frac{1}{T}$
FM3	$-\frac{1}{T} <$ Bias < 0
Pooled IV	Bias ≈ 0

approach, the size of the bias depends heavily upon the data generating process. In contrast, the bias in the adjusted Fama Macbeth methods *FM2* and *FM3* is hardly influenced by the true data generating process, but depends heavily on the number of periods, T, used to construct the average return $\bar{r}_i$. For simplicity, we have assumed that T is the same for all funds, but in reality the sample of funds is typically unbalanced with an increasingly small number of observations for earlier periods. In that case, the absolute bias in the *FM2* method is some weighted average of $\frac{1}{T_i}$, T_i being the number of periods available for fund i, which may be substantially larger than $\frac{1}{T}$, where T denotes the maximum number of sample periods. We shall illustrate this in the simulation exercise in the next section.

It is clear from all expressions above that the biases disappear if T tends to infinity, except for the standard *FM* method. With increasing T, the correlation between the estimation error in $M_{t-1}(r_{it})$ and any historical return (i.e. $r_{i,t-1}$) tends to zero. In practice, however, only a finite history is available for each fund in the sample such that the bias may not be negligible, particularly given the order of magnitude of persistence coefficients found in the literature. Moreover, as under the null hypothesis of no predictability in returns, the returns are uncorrelated over time, the bias is similar for all coefficients if additional lags are included in the regression ($J > 1$). So the cumulative bias in a regression with 8 lags included is of the order 8/T. This will be one of the points we will illustrate in the simulation exercise in the next section.

5.4 Properties of the Estimators: Numerical Results

To simplify the analytical derivations, we assumed that there was only one lag ($J = 1$) included in the regressions. To illustrate the numerical magnitude of the biases in some of the estimation methods when more than one lag is included, we performed a number of Monte Carlo simulation experiments. For the first experiment, we assume that true fund returns can be described by a one factor model with an unpredictable factor. This corresponds to a null hypothesis of no predictability in returns. In a second experiment, we examine the behavior of the five estimation methods when true funds returns do have a predictable component.

For the first experiment, we generate returns for a sample of 750 mutual funds over 60 periods. To do this, we follow the set-up of Brown, Goetzmann, Ibbotson and Ross [1992], whose parameter values were based on Ibbotson and Sinquefeld [1990], while increasing the frequency to quarterly observations. Quarterly returns are generated from the one factor model

$$r_{it} = \beta_i(R_{mt} - r_f) + u_{it}, \tag{5.20}$$

where the quarterly risk free rate r_f is taken to be 0.0175 (corresponding to an annual rate of 7%) and the quarterly risk premium $R_{mt} - r_f$ is assumed to be normal with mean 0.022 and standard deviation 0.104. The idiosyncratic error term u_{it} is independent of the risk premium $R_{mt} - r_f$, and also assumed to be normal with mean zero and variance σ_i^2, given by

$$\sigma_i^2 = k(1 - \beta_i)^2. \tag{5.21}$$

This relationship is a rough approximation to the relationship between nonsystematic risk and β that is often observed in mutual funds data. The value of k in our experiment equals 0.01337. Finally, the distribution of fund betas is assumed to be normal with mean 0.95 and a standard deviation of 0.25.

In the Monte Carlo experiment we generate 2500 samples with 750 funds observed over 60 consecutive quarters. Following Hendricks, Patel and Zeckhauser [1993], we now include eight lags in the regressions ($J = 8$). For the standard *FM* estimation method and the adjusted *FM2* and *FM3* methods, we estimate, for each sample, 52 cross-sectional regressions and computed the average coefficient estimates. For the adjusted *FM1* method only 44 cross-sectional regressions are performed. The *pooled IV* estimation method implies that only one regression has to be estimated for each sample. The numbers reported correspond to the average estimates of the 2500 replications and the average t-values.

For the first method, *FM*, we can expect a (small) bias due to the cross-sectional variation in expected returns. The second method, *FM1*, replicates Jegadeesh's solution by subtracting the average return over the eight quarters following[7] quarter t, which should yield unbiased estimates. The next two choices correspond to *FM2* and *FM3* and subtract the average return over the whole sample period and the predicted return from a CAPM time-series regression, respectively. Both methods are expected to yield a negative bias. The final method, *pooled IV*, is based on instrumental variables estimation of a pooled regression in terms of first differences of returns, and should yield unbiased results.

Table 5.2: **Average estimates and t-values; simulated data without persistence.** The table reports time-averages of the slope coefficient estimates obtained with cross-sectional regressions estimated by OLS for each period (t=9,...,60). For the methods adjusted FM1, FM2 and FM3, the dependent variable is in excess of an estimate of the expected return. All the numbers are averages over 2500 Monte Carlo replications. In each period, the full sample of 750 funds is available.

Average estimates (x 100), t-values in parentheses					
Estimation method	Standard FM	Adjusted FM1	Adjusted FM2	Adjusted FM3	Pooled IV
Dependent variable	r_{it}	$r_{it} - \bar{r}_i^*$	$r_{it} - \bar{r}_i$	$r_{it} - \hat{M}_{t-1}$	$r_{it}^* - r_{it-1}^*$
$\bar{\gamma}_1$	0.41	-0.00	-1.63	-1.45	0.39
	(0.23)	(-0.00)	(-0.95)	(-1.80)	(0.04)
$\bar{\gamma}_2$	0.50	0.06	-1.55	-1.46	-0.18
	(0.27)	(0.03)	(-0.91)	(-1.82)	(-0.02)
$\bar{\gamma}_3$	0.46	0.00	-1.59	-1.44	0.25
	(0.25)	(0.01)	(0.92)	(-1.81)	(0.03)
$\bar{\gamma}_4$	0.40	-0.06	-1.65	-1.46	0.08
	(0.22)	(-0.03)	(-0.95)	(-1.82)	(0.01)
$\bar{\gamma}_5$	0.43	-0.02	-1.62	-1.43	-0.16
	(0.24)	(-0.00)	(-0.94)	(-1.79)	(-0.00)
$\bar{\gamma}_6$	0.45	-0.00	-1.60	-1.43	-0.22
	(0.26)	(0.00)	(-0.93)	(-1.79)	(-0.04)
$\bar{\gamma}_7$	0.41	-0.04	-1.64	-1.45	0.17
	(0.23)	(-0.02)	(-0.95)	(-1.81)	(0.00)
$\bar{\gamma}_8$	0.51	0.09	-1.54	-1.42	-0.26
	(0.28)	(0.05)	(-0.89)	(-1.79)	(-0.04)
$\sum \bar{\gamma}_j$	3.57	0.03	-12.82	-12.97	0.07
	(0.51)	(0.09)	(-2.59)	(-5.08)	(0.09)

The results are summarized in Table 5.2. Clearly, the magnitude of the biases found corresponds closely to the analytical expressions given above. For the standard *FM* method, a small positive bias is found of approximately 0.004 in all slope coefficients, while for Jegadeesh's solution biases are negligible[8]. For adjusted Fama-MacBeth procedures *FM2* and *FM3*, corresponding to the Hendricks, Patel and Zeckhauser [1993] choices, a negative bias is found in

[7] Due to this choice, for the last eight quarters of data no cross-sectional regression can be performed; the estimates presented are averages over 44 quarters.

[8] That is, insignificantly different from zero, using the Monte Carlo standard errors

all slope coefficient estimates of the order of -0.016 and -0.014, respectively. Note that in the *FM3* approach, the market model used to estimate M_{t-1} corresponds to the true data generating process and will probably result in a better fit than commonly found in applied work. Although the negative numbers found seem small, the bias is shared by all coefficients such that the cumulative of all eight coefficients is biased by about -0.13. Interpreting this along the lines of Hendricks, Patel and Zeckhauser [1993], this implies that in the wake of a 1% superior performance, the cumulative residual loss is about 13 basis points over the next eight quarters[9]. Moreover, increasing the number of lags in the regression, would result in even more coefficients that are biased in the same direction. The *pooled IV* method gives coefficients that vary between -0.003 and +0.004 with rather high standard errors.

The average t-values reported in the table, except those for the *pooled IV* method, are based on the usual Fama-Macbeth standard errors and are thus adjusted for heteroskedasticity over time and over the funds. Compared to the other alternative Fama Macbeth approaches, the standard errors for the case with residual returns from the market model (*FM3*) are small. This is probably due to the fact that the variation over time in residual returns $(r_{it} - \hat{r}_{it})$ is much smaller than the variation in excess returns $(r_{it} - \bar{r}_i)$. Also note that the market model used in this approach is correctly specified by construction. While this will hardly affect the average coefficient estimates, it will reduce their variation over time. For the *pooled IV* method, t-values are calculated assuming homoskedasticity across time (but not across funds) and allowing for first order (moving average) autocorrelation in the differenced errors. The standard errors are substantially higher than for the other approaches. Apparently, robustness pays a price in terms of efficiency.

While for the adjusted *FM2* and *FM3* methods none of the slope coefficients is individually significantly different from zero (according to the average t-values), a joint test leads to rejection. Moreover, the cumulative residual gain, as measured by the estimates of $\sum_j \gamma_j$, is significantly different from zero for each of the biased methods *FM2* and *FM3*.

It is clear that the estimation error in estimating "equilibrium" returns induces a bias in the slope coefficient estimates, which in itself may be small, but may seriously affect economic conclusions. The biases are all negative, implying that it cannot induce spurious findings of "hot hands" in mutual funds. It may, however, indicate that the "hot hands" phenomenon is even stronger than reported by Hendricks, Patel and Zeckhauser [1993].

---

9 HPZ report a cumulative residual gain of 20 basis points over 8 quarters.

The set-up of the second experiment is comparable with the first one. However, quarterly returns are now generated by

$$r_{it} = \beta_i(R_{mt} - r_f) + \gamma_1 r_{i,t-1} + u_{it}, \tag{5.22}$$

with $\gamma_1 = 0.05$, while the other parameter values are left unchanged.[10] Essentially, the data generating process (5.22) includes a simple predictable pattern of past returns. Ideally, the estimation methods should yield a positive (and significant) coefficient for the first lag and zero values for the others. The results of 2500 Monte Carlo simulations are summarized in Table 5.3.

Table 5.3: **Average estimates and t-values; simulated data with first order persistence.** The table reports time-averages of the slope coefficient estimates obtained with cross-sectional regressions estimated by OLS for each period (t=9,...,60). For the methods adjusted FM1, FM2 and FM3, the dependent variable is in excess of an estimate of the expected return. All the numbers are averages over 2500 Monte Carlo replications. In each period, the full sample of 750 funds is available.

Average estimates (x 100), t-values in parentheses					
Estimation method	Standard FM	Adjusted FM1	Adjusted FM2	Adjusted FM3	Pooled IV
Dependent variable	r_{it}	$r_{it} - \bar{r}_i^*$	$r_{it} - \bar{r}_i$	$r_{it} - \hat{M}_{t-1}$	$r_{it}^* - r_{it-1}^*$
$\bar{\gamma}_1$	5.51	5.03	3.36	3.38	5.25
	(3.12)	(2.53)	(1.95)	(4.21)	(0.48)
$\bar{\gamma}_2$	0.40	-0.04	-1.64	-1.48	-0.22
	(0.23)	(-0.02)	(-0.95)	(-1.84)	(-0.02)
$\bar{\gamma}_3$	0.49	0.09	-1.55)	-1.44	0.01
	(0.28)	(0.04)	(-0.90	(-1.79)	(0.00)
$\bar{\gamma}_4$	0.36	-0.07	-1.67	-1.49	-0.26
	(0.20)	(-0.03)	(-0.96)	(-1.85)	(-0.02)
$\bar{\gamma}_5$	0.40	-0.01	-1.63	-1.47	0.40
	(0.23)	(-0.00)	(-0.95)	(-1.83)	(0.03)
$\bar{\gamma}_6$	0.42	0.01	-1.62	-1.46	-0.32
	(0.24)	(0.01)	(-0.94)	(-1.83)	(-0.04)
$\bar{\gamma}_7$	0.37	-0.09	-1.67	-1.48	0.15
	(0.20)	(-0.04)	(-0.97)	(-1.85)	(0.00)
$\bar{\gamma}_8$	0.41	-0.04	-1.74	-1.55	-0.29
	(0.23)	(-0.02)	(-1.00)	(-1.93)	(-0.04)
$\sum \bar{\gamma}_j$	8.36	4.88	-8.16	-6.99	4.72
	(1.55)	(0.87)	(-1.71)	(-3.25)	(0.07)

The order of the biases, present in the five estimation methods, in the case that the true data generating process contains a predictable component are comparable to those found with the unpredictable process. The *FM2* and *FM3* methods seriously underestimate the true coefficients. Jegadeesh's approach, *FM1*, produces estimates close to the coefficient $\gamma_1 = 0.05$, but as mentioned before, has the disadvantage that future returns are required. Despite the fact that it uses a longer sample period, the standard errors of the *pooled IV* approach are approxi-

[10] Note that this increases the overall average excess return by about 5%.

mately five times as large as those of the *FM1* method, which seems to make the IV approach inappropriate for applied work.

Until now, our sample of mutual funds was not very representative for samples used in empirical work, as it is assumed that fund returns are available over the whole sample period of 60 quarters. To see how the conclusions are affected if funds returns are observed over a limited history only, we took our previous sample of the first experiment and, going back in time, *randomly* removed 2% of the funds in each quarter. This results in an average number of funds in the first quarter of 223 (30%), which seems reasonable given the growth in the number of U.S. mutual funds over the last decade. It is important to realize two things. First, it is not the number of funds that is relevant for the biases, but the (average) number of periods used to estimate M_{t-1}. Using the 223 funds existing over the whole sample period would produce biases similar to those reported in Table 5.2. Second, there is no survivorship bias here, as the disappearing funds are selected completely randomly. Any survivorship bias would come above the estimation bias discussed here (although the two biases can have opposite sign and may cancel out).

Table 5.4: **Average estimates and t-values; selected sample from simulated data.** The table reports time-averages of the slope coefficient estimates obtained with cross-sectional regressions estimated by OLS for each period (t=9,...,60). For the methods adjusted FM1, FM2 and FM3, the dependent variable is in excess of an estimate of the expected return. All the numbers are averages over 2500 Monte Carlo replications. In each period, a random 2% of new funds are added to the sample, such that in the last period 750 mutual funds exist.

Average estimates (x 100), t-values in parentheses					
Estimation method	Standard FM	Adjusted FM1	Adjusted FM2	Adjusted FM3	Pooled IV
Dependent variable	r_{it}	$r_{it} - \bar{r}_i^*$	$r_{it} - \bar{r}_i$	$r_{it} - \hat{M}_{t-1}$	$r_{it}^* - r_{it-1}^*$
$\bar{\gamma}_1$	0.37	-0.10	-2.08	-1.80	0.21
	(0.18)	(-0.05)	(-1.07)	(-1.57)	(0.07)
$\bar{\gamma}_2$	0.45	-0.00	-2.00	-1.78	-0.16
	(0.23)	(-0.00)	(-1.03)	(-1.55)	(-0.06)
$\bar{\gamma}_3$	0.51	0.06	-1.94	-1.78	0.23
	(0.26)	(0.03)	(-1.00)	(-1.55)	(0.06)
$\bar{\gamma}_4$	0.44	-0.03	-2.01	-1.79	-0.34
	(0.22)	(-0.01)	(-1.04)	(-1.56)	(-0.05)
$\bar{\gamma}_5$	0.48	0.03	-1.97	-1.76	0.30
	(0.24)	(0.01)	(-1.02)	(-1.54)	(0.05)
$\bar{\gamma}_6$	0.45	-0.02	-2.00	-1.83	0.02
	(0.22)	(-0.01)	(-1.03)	(-1.59)	(0.03)
$\bar{\gamma}_7$	0.42	-0.06	-2.03	-1.79	0.25
	(0.20)	(-0.03)	(-1.05)	(-1.57)	(0.04)
$\bar{\gamma}_8$	0.42	-0.05	-2.03	-1.80	-0.07
	(0.21)	(-0.02)	(-1.04)	(-1.57)	(-0.05)
$\sum \bar{\gamma}_j$	3.54	-0.17	-16.06	-14.33	0.44
	(0.52)	(-0.06)	(-2.83)	(-4.39)	(0.09)

The results for the selected sample are presented in Table 5.4. As expected, the biases increase in absolute size compared to those reported in Table 5.2, except for the standard *FM* approach, the adjusted *FM1* approach and the *pooled IV* method.

Finally, we checked how sensitive the reported biases were for the particular parameter values chosen. As expected, varying the parameter values within reasonable bounds hardly had an effect on the numbers in the Tables 5.2, 5.3 and 5.4, except for the standard *FM* estimator. The t-values appeared less insensitive to the parameter values; in particular a smaller variance of the market risk premium led to an increase of the t-values for all estimators, except for the *FM3* approach. Substantial changes, however, were encountered when the number of periods was reduced to 30, in which case the biases almost doubled. It is important to keep this in mind as an analysis based on yearly data would produce similar results if the number of years employed in estimating M_{t-1} is the same as the number of quarters used in this study. Clearly, 60 years of data are available for only very few mutual funds, so that with annual data the biases encountered may be much larger than those reported here.

5.5 Empirical results for 1986-1994

Several recent empirical studies report short-run persistence in mutual fund performance. In light of our results of the previous two sections, we shall, in this section, empirically examine whether mutual funds do have a pattern of predictable returns using a sample of U.S. open-end mutual funds over the period 1986-1994. This analysis will illustrate the order of magnitudes of persistence coefficients that are relevant for applied work, so as to clarify the importance of the, seemingly small, biases reported in the previous sections.

Hendricks, Patel and Zeckhauser [1993], looking at short-run persistence of mutual fund returns over the period 1974-1988, found a pattern of positive coefficients for the first four lagged quarterly returns, while lags 5 to 7 were negative, and lag 8 was positive again. The cumulative gain in expected returns by selecting the funds that have an above average return is, according to their estimates, about 30 basis points over the next four quarters, but declines to about 20 basis points after eight quarters. Malkiel [1995], using data for 1971-1991, reports strong persistence for the 1970s but argues that this is largely gone in the later part of the sample. Given these findings and given the substantial growth in the mutual funds' universe over the last decade, it is a reasonable strategy for this chapter to focus on a relatively short but recent nine-year period.

To examine whether a pattern similar to the one found by Hendricks, Patel and Zeckhauser [1993] holds over the period 1986-1994, we employ two samples of quarterly mutual fund returns taken from the Morningstar Mutual Fund Database. Morningstar reports information about all open-end mutual funds on a monthly basis. We converted monthly returns to quarterly figures and included the fund returns until the moment of disappearance. This does not, however, eliminate potential survivorship biases as those reported in Brown, Goetzmann, Ibbotson and Ross [1992] and Hendricks, Patel and Zeckhauser [1997], but it is close to existing studies of performance persistence (including Hendricks, Patel and Zeckhauser [1993] and Carhart [1997a]).

The basic sample includes funds that meet the following selection criteria. First, the fund has an observation record of at least nine quarters[11]. Second, funds that invest more than 50% in bonds, but nevertheless advertise as "equity fund" are excluded from the sample. As a consequence of our first criterium, funds that ceased to exist before January 1988 are also excluded from the sample. The resulting sample varies from 711 mutual funds in the first quarter of 1986 to 1422 funds in the fourth quarter of 1994.

Following several papers in the area, our second sample contains a relatively homogeneous sample of equity funds, selected out of the basic sample, with as investment objective growth stocks. The size of this sample varies between 171 funds in the first quarter of 1986 and 353 mutual funds in the fourth quarter of 1994. For both samples we assume that all dividends are reinvested in the mutual fund at the end of the quarter in which the dividends are distributed. For the riskless rate we take the quarterly return on one-month U.S. Treasury bills, collected from the Ibbotson Index database. In order to apply the *FM3* method, we use the Standard & Poor 500 index, also collected from the Ibbotson Index database, as the market return. All returns are net of transaction costs, fees, and expenses, but are gross for any sales charges.

Although the choice of the number of lags in the regressions is a bit arbitrary, we follow Hendricks, Patel and Zeckhauser [1993], and only include up to eight lags, as in our simulation experiments. Including more lags would enable estimation of additional medium-run persistence effects, but effectively reduces the number of observations in estimation. The estimation results of the five methods are summarized in Table 5.5, for the basic sample, and Table 5.6 for the sample of growth stocks.

As discussed above, the *FM2* method has an expected bias that is a weighted average of $-\frac{1}{T_i}$, where T_i is the number of periods available for fund i. In this empirical study, the maximum number of observations available is 36, which means that we can expect a bias of at least

[11] To apply *FM2*, we also need eight future observations, so that for this approach an observation history of at least 17 periods is required. This leads to a sample of 1209 funds (284 growth funds) in the last quarter of 1992.

Table 5.5: **Persistence estimates and t-values: basic sample 1986-1994.** The table reports the estimated persistence coefficients using the basic sample of mutual funds. The estimates of the adjusted FM1 method are based on cross-sectional regressions, estimated by OLS, for each quarter in 1988 through 1992 (t=9,..,28). In contrast, the standard FM and adjusted FM2 and FM3 approaches are based on 28 cross-sectional regressions. The estimates reported are the time-averages of the slope coefficient estimates. For the methods adjusted FM 1,2 and 3, the dependent variable is in excess of an estimate of the expected return. The numbers in parentheses are t-values.

Estimates (x 100) of Persistence Coefficients					
Estimation method	Standard FM	Adjusted FM1	Adjusted FM2	Adjusted FM3	Pooled IV
Dependent variable	r_{it}	$r_{it} - \bar{r}_i^*$	$r_{it} - \bar{r}_i$	$r_{it} - \hat{M}_{t-1}$	$r_{it}^* - r_{it-1}^*$
$\bar{\gamma}_1$	1.60	-3.86	-3.67	-1.94	4.54
	(0.32)	(-0.64)	(-0.75)	(-0.48)	(0.62)
$\bar{\gamma}_2$	1.16	-3.18	-3.42	-0.78	-0.31
	(0.26)	(-0.67)	(-0.80)	(-0.21)	(-0.39)
$\bar{\gamma}_3$	17.43	19.57	12.81	9.79	33.43
	(4.25)	(3.49)	(3.26)	(3.03)	(4.18)
$\bar{\gamma}_4$	0.11	4.90	-3.59	-1.98	3.65
	(0.03)	(0.98)	(-0.96)	(-0.58)	(0.58)
$\bar{\gamma}_5$	-7.39	-6.18	-11.03	-10.32	-0.42
	(-1.70)	(-1.40)	(-2.52)	(-2.21)	(-0.07)
$\bar{\gamma}_6$	-1.08	0.01	-5.49	-4.48	0.23
	(-0.38)	(0.00)	(-1.93)	(-1.51)	(0.03)
$\bar{\gamma}_7$	0.49	-4.96	-4.21	-4.13	12.87
	(0.15)	(-1.60)	(-1.33)	(-1.25)	(1.65)
$\bar{\gamma}_8$	12.63	9.20	7.98	3.22	17.72
	(4.19)	(2.26)	(2.56)	(1.15)	(2.51)
$\sum \bar{\gamma}_j$	24.94	15.40	-10.60	-10.62	71.71
	(3.03)	(1.60)	(-1.48)	(-1.48)	(1.58)

-0.028 in the estimated coefficients of this method. According to our simulation experiments, the bias in the estimates of the *FM1* method is negligible. Note, however, that in our case the estimates are the averages of only 20 cross-sectional regressions due to the fact that the unconditional expectation M_{t-1} is estimated from eight future observations. In contrast, the standard *FM* estimates are based on 28 cross-sectional regressions. Although the exact size of the bias present in the latter estimation method is dependent on the true data generating process, we expect a positive sign. This suggests that the true persistence coefficients are somewhat smaller than the estimates of the standard Fama Macbeth approach. The estimates of the *pooled IV* approach differ substantially from the estimates of the standard *FM* and *FM1* method. As already suggested, the *pooled IV* method suffers from large standard errors, which makes this approach less suitable for applied work.

Looking at the estimates of the adjusted *FM1* method, there appears to be some evidence of persistence in the basic sample of mutual funds, but the pattern is rather erratic. Given the accuracy of the individual estimates, it does not seem advisable to develop a dynamic buy-and-sell strategy from these numbers. A strategy that selects funds with a 1% superior performance,

Table 5.6: **Persistence estimates and t-values for selected sample of growth funds.** The table reports the estimated persistence coefficients using a sample of mutual funds with investments style 'growth'. The estimates of the adjusted FM1 method are based on cross-sectional regressions, estimated by OLS, for each quarter in 1988 through 1992 (t=9,..,28). In contrast, the standard FM and adjusted FM2 and FM3 approaches are based on 28 cross-sectional regressions. The estimates reported are the time-averages of the slope coefficient estimates. For the methods adjusted FM 1,2 and 3, the dependent variable is in excess of an estimate of the expected return. The numbers in parentheses are t-values.

Estimates (x 100) of Persistence Coefficients,					
Estimation method	Standard FM	Adjusted FM1	Adjusted FM2	Adjusted FM3	Pooled IV
Dependent variable	r_{it}	$r_{it} - \bar{r}_i^*$	$r_{it} - \bar{r}_i$	$r_{it} - \hat{M}_{t-1}$	$r_{it}^* - r_{it-1}^*$
$\bar{\gamma}_1$	4.05	0.03	-0.84	-0.76	2.18
	(0.72)	(0.00)	(-0.15)	(-0.16)	(0.27)
$\bar{\gamma}_2$	5.62	1.68	1.61	3.83	4.82
	(1.11)	(0.27)	(0.32)	(0.84)	(0.54)
$\bar{\gamma}_3$	12.93	11.62	8.92	6.56	43.65
	(2.40)	(1.62)	(1.73)	(1.49)	(4.77)
$\bar{\gamma}_4$	3.10	6.27	-4.88	-0.31	2.86
	(0.79)	(1.40)	(-0.13)	(-0.08)	(0.42)
$\bar{\gamma}_5$	-8.03	-2.34	-11.08	-10.89	2.53
	(-1.52)	(-0.50)	(-2.10)	(-1.96)	(0.36)
$\bar{\gamma}_6$	-2.80	0.82	-3.63	-3.22	-3.34
	(-0.08)	(0.19)	(-1.12)	(-0.93)	(-0.44)
$\bar{\gamma}_7$	-1.81	-3.65	-5.26	-5.46	19.86
	(-0.55)	(-1.07)	(-1.61)	(-1.58)	(2.10)
$\bar{\gamma}_8$	8.84	8.73	5.20	1.95	17.79
	(3.17)	(2.11)	(1.67)	(0.73)	(2.11)
$\sum \bar{\gamma}_j$	21.90	23.20	-9.90	-8.30	90.53
	(2.24)	(1.79)	(-0.52)	(-0.75)	(1.81)

and keeps these in portfolio for eight consecutive quarters, leads to an expected cumulative residual gain of 0.15%, with a standard error of 0.10%. The conclusions from the inconsistent *FM2* and *FM3* methods, on the other hand, would be substantially different with a cumulative residual *loss* of 0.10% and a standard error of 0.07%. The estimates using the standard *FM* approach, reported in the first column, seems to be upward biased, as can be expected from the analytical and Monte Carlo results, while the *pooled IV* estimates in column 5 produces substantially different results, with substantially higher standard errors. Most methods seem to have in common that lags 3 and 8 are important with significantly positive coefficients.

For the more homogenous subsample of growth funds, our *FM1* results, reported in column 2 of Table 5.6, show a pattern of persistence that corresponds fairly closely to the one reported by Hendricks, Patel and Zeckhauser [1993] for the period 1974 to 1988. Note, however, that the latter results were based on the adjusted methods *FM2* and *FM3*, which - in our case - would yield substantially different outcomes. Apparently, an investment strategy based on selecting growth-oriented mutual funds with an above average performance over the last four quarters still proves valuable for the period 1986-1994. The estimated cumulative gain in expected returns by selecting funds with a relatively high return compared to other growth funds is about 23 basis points over the next eight quarters. The associated standard error, however, corresponds to 13 basis points. Again, note that the conclusions from the adjusted *FM2* and *FM3* approaches would be substantially different with a cumulative loss of approximately 9 basis points. Recall that this is a biased estimate and does not represent the actual expected gain or loss from the above-mentioned strategy.

5.6 Concluding Remarks

In this chapter we examined a number of estimation methods used to detect patterns of predictable returns. As expected returns vary over the funds, most of these methods employ some estimate of these expected returns to prevent the problem of cross-sectional correlation, as discussed by Jegadeesh [1990]. Our analytical results show that estimation errors in the expected returns may induce a spurious pattern of short-run persistence. The bias in the persistence coefficients is, on average, close to $-\frac{1}{T}$, where T is the number of periods used to estimate the expected returns. As this bias hardly depends on the true data generating process, this result is of particular concern when using lower frequency data, where only a limited number of time series observations is available. As an illustration, we considered the approaches taken in Hendricks, Patel and Zeckhauser [1993], which had biases in each slope coefficient of approximately -0.02, corresponding to a cumulative bias (over eight lags) of -0.16.

Jegadeesh's [1990] approach to eliminate such biases requires estimation of expected returns over future observations, instead of past returns. Although this methods leads to unbiased estimates, the approach has as a disadvantage that the most recent observation periods are actually not used in the estimation of the short-run persistence coefficients. This is particularly cumbersome if time-variation in these coefficients can be expected. As an alternative, we suggest another estimation approach, which corresponds to instrumental variables estimation of the model in first differences, using the pooled data. Unfortunately, this approach, based on the elimination of fixed individual effects in dynamic panel data models, is, though consistent, rather inefficient, such that accurate statements about the true persistence coefficients are hard to make.

The second part of the chapter empirically examined the short-run persistence in a sample of equity funds and a subsample of growth equity funds, over the period 1986-1994. The results show that an investment strategy based on identifying the winning growth-oriented mutual funds increases the expected return on a portfolio of mutual funds. In particular, a strategy of selecting every quarter the funds with high returns, relative to other funds, over the last four quarters, can significantly increase the expected return. Although the estimates of Hendricks, Patel and Zeckhauser [1993] over the period 1975 to 1988 were negatively biased, they found a similar pattern. Apparently, the hot hands phenomenon reported by these authors still exists in the period 1986 to 1994. It must be stressed, however, that the estimates of the individual coefficients are not very accurate and, moreover, the results are quite sensitive to the estimation method employed. At the least, this implies that the development of dynamic trading strategies from these results is a dangerous exercise.

Chapter 6

Eliminating Biases in Evaluating Mutual Fund Performance from a Survivorship Free Sample

Poor performing mutual funds are less likely to be observed in the data sets that are typically provided by data providers. This so-called survivor problem can induce a substantial bias in measures of the performance of the funds and the persistence of this performance. Many studies have recently argued that survivorship bias can be avoided by analyzing a sample that contains returns on each fund up to the period of disappearance using standard techniques. Such data sets are usually referred to as 'survivorship free'. In this chapter we show that the use of standard methods of analysis on a 'survivorship free' data-set typically still suffers from a bias and we show how one can easily correct for this using weights based on probit regressions.

Using a sample with quarterly returns on U.S. based equity funds, we first of all model how survival probabilities depend upon historical returns, the age of the fund and upon aggregate economy-wide shocks. Subsequently we employ a Monte Carlo study to analyze the size and shape of the survivorship bias in various performance measures that arise when a 'survivorship free database' is used with standard techniques. In particular, we show that survivorship bias induces a spurious U-shape pattern in performance persistence. Finally, we show how a weighting procedure based upon probit regressions can be used to correct for the bias. In this way, we obtain bias-corrected estimates of abnormal performance relative to a one-factor and the Carhart [1997a] four-factor model, as well as its persistence. Our results are in accordance with the persistence pattern found by Carhart [1997a], and do not support the existence of a hot hand phenomenon in mutual fund performance.

6.1 Introduction

Many empirical studies in finance potentially suffer from survivorship bias. This point has recently been stressed by e.g. Brown, Goetzmann and Ross [1995] and Carhart [1997b]. In this chapter we focus on the impact of survivorship bias in measuring mutual fund performance. Poor performing mutual funds are less likely to be observed in the data sets that are typically available. This so-called survivor problem can induce a substantial bias in measures of the performance of the funds and the persistence of this performance. Many studies (see, e.g. Grinblatt and Titman [1989a], Brown and Goetzmann [1995], Malkiel [1995] and Wermers [1997]) have recently argued that survivorship bias can be avoided by using standard techniques on a sample that contains returns on each fund up to the period of disappearance. Such data-sets are usually referred to as 'survivorship free'. As stressed by Carhart [1997b], the analysis of a 'survivorship free' database with standard techniques will in general still yield biased estimates of performance measures, because poor performing funds are underrepresented. Carhart [1997b] refers to this bias as a 'look ahead bias'. In this chapter we analyze the relative size of these biases for U.S. based equity funds and show how these biases can be eliminated using appropriate correction methods.

Empirical studies by Brown and Goetzmann [1995], and Elton, Gruber and Blake [1996] indicate, as may be expected, that a bad record of returns is one of the main reasons for fund disappearance. If this is the case, a simple analysis of average returns for a sample of mutual funds still in existence at the end of the sample period tends to be upward biased due to the relative absence of low returns. Because of cross-sectional variation in expected returns, this bias does not necessarily disappear in a survivorship free sample. It is sometimes claimed (Grinblatt and Titman [1989a], Blake, Elton and Gruber [1993]) that the effect of survivorship bias on average returns is between 0.1% and 0.4% per year, although the implicit underlying assumptions on the survival process are not clear. For alternative and more sophisticated measures of performance, and its persistence, survivorship can lead to a wide range of spurious empirical regularities, the form of which will depend upon the survival process (see, e.g. Brown et al. [1992], Brown, Goetzmann and Ross [1995] and Hendricks, Patel and Zeckhauser [1997]).

In this chapter we empirically study the performance of U.S. based open-end mutual funds for the period 1989-1995, explicitly taking into account the problem of survivorship bias. Following the micro-economics literature on sample selection (starting with Heckman [1976, 1979]), we model the process that determines attrition from the sample, and subsequently analyze it jointly with (the underlying model of) performance evaluation. As a consequence, the goal of this chapter is threefold. First, we determine the factors that affect a fund's probability to close or merge and leave the sample. The longitudinal probit model that we propose extends the model in Brown and Goetzmann [1995] by allowing for aggregate macro-economic shocks. Second, we analyze the effects of this survival process on a range of performance measures using a Monte Carlo experiment and, third, we show how one can correct for these survivorship biases and apply this to the sample of equity funds. Our results show that historical returns are an important determinant for fund survival, that survivorship bias in performance measurement can be substantial, and that knowledge of the survival process enables fairly simple corrections for survivorship biases. Using Carhart's [1997a] four-factor model for evaluating mutual fund performance, we find a persistence pattern that is similar to the one reported in Carhart [1997a], although the latter may be subject to bias.

The remainder of this chapter is organized as follows. A stylized example in Section 6.2 illustrates the potential problem of survivorship bias in performance measurement using a survivorship free as well as a survivors-only sample of mutual funds. In Section 6.3 we describe the sample of U.S. based mutual funds that we employ. We show that the total number of funds that leaves the sample is substantial and their average return is substantially less than for surviving funds. This indicates the potential for survivorship bias and the need to correct for it. In Section 6.4 we model survival probabilities and examine factors that determine fund disappearance. We also analyze whether funds with different investment objectives, such as growth stocks or foreign stocks, have different probabilities of survival. A Monte Carlo study, presented in Section 6.5, shows the effect of survivorship bias on various methods analyzing mutual fund performance. The empirical survival process, in which historical returns over at most three years play a role, induces a spurious pattern of performance persistence that is U-shaped, rather than J-shaped as in Hendricks, Patel and Zeckhauser [1997]. In Section 6.6 we show how the survival model can be used to construct weights that can be applied to correct for survivorship biases. The empirical implementation of this approach is presented in Section 6.7, where we examine persistence in the performance of U.S. open-end equity funds over the period 1989-1994, using a simple one-factor model and Carhart [1997a]'s four-factor model. By and large, our results support the findings of Carhart [1997a] and do not indicate the existence of a hot hands phenomenon in mutual fund performance. Section 6.8 summarizes the main results and presents some concluding remarks.

6.2 A Stylized Example

In order to show that the use of a survivorship free sample to evaluate mutual fund performance still yields biased estimates, we will in this section analyze a simple example that illustrates the causes and sizes of the impact that survivorship effects can have on performance measures. More explicitly, we will examine the effect on the average return of a mutual fund or a sample of funds given that mutual fund survival depends on past returns.

Assume that a population of mutual funds exists, indexed $i = 1, .., M$. In two consecutive sample periods, each of them realizes a return $r_{it} = \mu^H$ with probability p_i, or $r_{it} = \mu^L$ with probability $1 - p_i$, where p_i comes from a cross-sectional distribution with mean p. Consequently, the expected return on a mutual fund is:

$$\mu_i = p_i\mu^H + (1 - p_i)\mu^L.$$

In the first sample period each fund is observed, while in the second period we observe a fund with probability one if it had a return $r_{it} = \mu^H$ in the previous period and with probability q if it had return $r_{it} = \mu^L$. Indexing data availability in the second period by $y_i = 1$, we thus have that

$$P\{y_i = 1|r_{i1}\} = q + (1 - q)\frac{r_{i1} - \mu^L}{\mu^H - \mu^L} \tag{6.1}$$

and

$$P\{y_i = 1\} = p_i + (1 - p_i)q. \tag{6.2}$$

The standard estimator for the expected return μ_i of fund i from a 'survivorship free' database

$$\hat{\mu}_i = \frac{r_{i1} + y_i r_{i2}}{1 + y_i}. \tag{6.3}$$

(Note that r_{i2} is missing if $y_i = 0$.) By using the probabilities for each of the possible outcomes, it is easily verified that this is not an unbiased estimator for μ_i. In particular, it can be shown that

$$E\{\hat{\mu}_i\} - \mu_i = [\frac{1}{2}p_i(1 - q)(1 - p_i)](\mu^L - \mu^H), \tag{6.4}$$

indicating that (6.3) underestimates the expected return μ_i. Consequently, even a survivorship free sample is not free of survivorship effects in the sense that the properties of standard estimators can be affected by the survival process. The bias increases as the probability of survival q decreases. Not surprisingly, the bias disappears if $\mu^H = \mu^L$, if the survival probability q equals one or if $p_i = 0$ or $p_i = 1$.

Conditional upon fund survival in period 2, the expected value of $\hat{\mu}_i = \frac{1}{2}(r_{i1} + r_{i2})$ is given by

$$E\{\hat{\mu}_i | y_i = 1\} - \mu_i = \frac{\frac{1}{2}p_i(1-q)(1-p_i)}{p_i + (1-p_i)q}(\mu^H - \mu^L), \tag{6.5}$$

which indicates that the estimator based upon the survivors-only sample overestimates the expected return μ_i. This is due to the relative absence of bad returns.

Intuitively, the fact that the estimator for the expected return conditional upon fund survival yields a positive bias, due to the relative absence of low returns, suggests that in order to obtain an unbiased estimate the weight for the observed low returns should be increased, while the observed high returns should have a lower weight. In Section 6.6 we show that using a weight factor that equals the inverse of the normalized probability that fund i is kept in the sample yields unbiased estimates. Consequently, to estimate the expected return μ_i for fund i, the appropriate weight is the ratio of the unconditional (6.2) and conditional survival probability (6.1) and is given by

$$w_i = \frac{p_i + (1-p_i)q}{q + (1-q)\frac{r_{i1} - \mu^L}{\mu^H - \mu^L}}. \tag{6.6}$$

It is easily verified that the adjusted estimator $\tilde{\mu}_i = \frac{1}{2}w_i(r_{i1} + r_{i2})$ is an unbiased estimator based upon the sample characterized by $y_i = 1$. Note that the weights w_i depend upon returns, r_{i1} in this case, and are thus endogenous. Moreover, they depend upon the unknown parameters p_i. This seems strange but is not a problem because the numerator of the weights is just $P\{y_i = 1\}$ which is directly identifiable from the data, using observations on other funds.

6.3 Stylized Facts on Survival of U.S. Equity Funds

In order to examine the importance of the effect of conditioning upon survival for empirical performance studies, we employ a data set selected from the Morningstar Mutual Funds Ondisc database (February 1995 edition). This database contains monthly information on more than 6000 U.S. based open-end equity as well as fixed income mutual funds. This sample suffers from survivorship because only funds that existed at the end of the sample period (February 1995) are included. Many mutual funds have disappeared from the sample because they have merged with other funds or are closed down completely. In the latter case it is possible that the management of the fund has decided to change to a closed-end fund, or that the investors of the fund are offered the opportunity to withdraw their money and invest it in another fund of the same investment company. A first step in obtaining results free of survivorship bias, is the inclusion of attrited funds in the sample. We did so by extending the database with the mutual funds that disappeared between the first month of 1989 and the last month of 1994. In the sequel we will refer to this data set, covering 1989-1994, as the combined 'survivorship free' sample[1].

In this chapter, following previous studies, we concentrate on equity funds. During the period January 1989 to December 1994, we observe 2678 funds with their name, objective, the year of fund inception, and monthly returns until the month of disappearance. In contrast to Carhart [1997b] we also included specialty funds (i.e. 273 sector funds), internationally diversified U.S. based funds (490) and a number of funds which advertise as 'balanced fund' and invest less than 50% in fixed income securities (180). For 79 funds Morningstar did not report an investment objective. For 33 funds of this group the investment objective was obvious from their name. The remaining 46 funds were classified as having an 'other' investment objective. Table 6.1 presents the number of funds by inception year and the number of them that did not survive the period January 1989 to December 1994, aggregated to a yearly level. Note that we aggregated all funds with an inception date before 1977, which explains the relative large number of 279 funds with inception year 1976.

Table 6.1 shows that 498 of the mutual funds in our sample disappeared between January 1989 and December 1994 due to merger or liquidation, which corresponds with a yearly average of 5.3%. This estimate differs from Carhart [1997b], who reports a non-survival rate of 3.6% over the period 1962 to 1995, which increases to 4.6% for the period 1989 to 1994. The remaining difference with our yearly average is due to inclusion of types of funds in our sample that have relatively low survival rates (compare Table 6.3 below). Furthermore, looking at, for instance, the 173 funds that started in 1990, 29.5% of them already disappeared within the next four years. A similar pattern seems to hold for other years, so that a first conclusion would be that a large part of the defunct funds disappeared at a relatively young age, age being defined as the time elapsed since fund inception. Apparently, it is not only the case that the number of mutual funds has grown at an increasing rate over the last decade, but also that the relative number of funds that has closed down or merged has increased significantly. At a more disaggregated level (not reported in the table), it appears that in some months relatively many mutual funds leave the sample while in other months almost no funds disappear, indicating that common aggregate factors may play a role in fund disappearance as well.

It is often claimed that a bad record of fund returns is one of the main reasons that funds disappear from a sample (see, e.g. Elton, Gruber and Blake [1996]). Low returns compared to

[1] Unfortunately, Morningstar was unable to provide information about funds that ceased to exist before 1989.

Table 6.1: **Number of Non-survivors.** The table reports the annual number of funds by inception year since 1976 and the annual number that ceased to exist between 1989 through 1994 . The row labelled 'Non-surv. Rate' contains the number of disappearing funds divided by the total number of funds at the beginning of the year.

Inception year	total in	out in year: 1989	1990	1991	1992	1993	1994	total out
≤1976	279	5	3	4	7	16	8	43
1977	14		1		1	2		4
1978	13			2	2			4
1979	11	1		1		1		3
1980	10					1		1
1981	26	2		1	1			4
1982	33	1		2	2	3	1	9
1983	58	4	3	3	3	3	2	18
1984	75		2	3	2	3	5	15
1985	98	2	5	5	3	6	4	25
1986	143	1	11	8	9	13	7	49
1987	162	3	6	9	7	16	11	52
1988	140	2	2	7	13	14	10	48
1989	107		3	5	4	12	7	31
1990	173			11	6	18	16	51
1991	187				8	7	23	38
1992	277					17	15	32
1993	550					13	40	53
1994	322						18	18
total:	2678	21	36	61	68	145	167	498
Non-surv. Rate (%/yr)		1.98	3.14	4.75	4.82	8.95	8.25	5.31

other funds as well as low returns relative to some benchmark portfolio seem to be a reason for the management of the fund to close down the fund or to let the fund merge (see Brown and Goetzmann [1995]). Table 6.2 presents the average quarterly returns for the period 1989-1994 for the funds that survived until the end of 1994, for the funds that did not survive, and for the combined sample. For comparison, we also added the quarterly returns on the Standard and Poor 500 and the returns on a three month Treasury Bill over the same period.

In accordance with other studies, like Malkiel [1995], it appears that in almost all quarters the surviving funds had a higher average return than the non-surviving funds, indicating that low returns increase the probability of disappearance. Furthermore, the average annual return over the period 1989 through 1994 for the sample containing the surviving funds is 0.64% higher than for the combined sample of funds, i.e. 11.44% versus 10.80%. In contrast, Malkiel [1995] finds a difference of 1.50% over the period 1982 through 1991, Elton, Gruber and Blake [1995] even find a difference of 1.87% for 1976 through 1993, while Brown and Goetzmann [1995] report a difference of 0.80% over the period 1977-1987. Note that all estimates for this effect of survivorship in computing average returns are higher than the at most 0.40%, that was claimed by Grinblatt and Titman [1989a]. In most quarters the average return of the mutual

Table 6.2: **Average quarterly returns.** The table reports the average quarterly returns for the funds that survived until the end of 1994, average quarterly returns for the funds that ceased to exist during 1989-1994, and the average return for the combined sample. Furthermore, we present the quarterly returns for the Standard and Poor 500 and Treasury bills. The columns labelled 'Number' contain the number of funds over which the average quarterly return is calculated.

quarter	Surviving Funds Mean return	Number	Non-Survivors Mean return	Number	Combined Mean return	return S&P500	return t-bill
1989/01	6.50	819	5.06	296	6.12	7.03	1.85
/02	6.36	830	5.28	294	6.08	8.80	2.19
/03	9.79	843	7.97	289	9.32	10.64	2.10
/04	0.34	858	-0.45	286	0.14	2.05	1.98
1990/01	-2.72	888	-2.78	298	-2.74	-3.02	1.79
/02	5.75	921	4.65	306	5.47	6.29	2.00
/03	-14.91	951	-13.35	300	-14.54	-13.78	1.95
/04	6.83	980	5.01	301	6.40	8.95	1.86
1991/01	15.16	1025	12.30	308	14.50	14.56	1.44
/02	-0.79	1058	-1.21	301	-0.88	-0.21	1.43
/03	6.91	1095	5.99	284	6.72	5.38	1.41
/04	7.24	1129	6.03	278	7.00	8.36	1.20
1992/01	-0.63	1184	-1.75	274	-0.84	-2.55	0.96
/02	-1.07	1213	-1.34	265	-1.12	1.97	0.92
/03	1.44	1298	0.93	254	1.36	3.10	0.83
/04	6.84	1374	5.63	242	6.66	5.10	0.75
1993/01	4.88	1515	3.59	241	4.70	4.28	0.71
/02	2.86	1603	2.10	253	2.75	0.51	0.71
/03	5.69	1732	4.51	164	5.59	2.56	0.75
/04	4.19	1871	2.36	150	4.05	2.31	0.70
1994/01	-3.13	2043	-3.46	136	-3.15	-3.82	0.73
/02	-2.03	2181	-2.61	115	-2.06	0.41	0.90
/03	5.69	2183	2.77	21	5.66	4.92	1.01
/04	-2.45	2184	.	.	-2.45	-0.03	1.20
Mean	2.86	.	.	.	2.70	3.08	1.31

funds underperforms the S&P500 which could be due to the fact that the equity funds hold bonds and liquidities as well. On an annual basis we find a difference between the S&P500 and the combined sample of funds of 1.52%, i.e. 12.32% versus 10.80%.

In order to examine whether the survival rate varies with the funds' investment objective, we broke down the sample by investment styles. While Morningstar reports investment objectives in thirteen different categories, we chose to follow other studies (Malkiel [1995], Brown and Goetzmann [1995]), and decided to split the sample, for ease of comparison, into six categories. The category 'other' represents the equity funds that could not be clearly assigned to any of the other five categories, so it contains, for instance, the funds that advertise as equity fund and invest less than 50% in fixed income securities as well as the funds with unknown investment objective. Table 6.3 presents the average quarterly return over the period 1989-1994 for all funds as well as for the subset of funds that survived until the end of 1994 for each of the six investment objectives[2]. It appears that the categories 'specialty' and 'other' had the highest

Table 6.3: **Summary Statistics by Objective Category.** The table shows for six investment objectives categories the average quarterly return for the combined sample and the average quarterly return for the funds that survived until the end of 1994 as well as the number of funds in each category, the number of non-survivors and the corresponding drop-out percentage.

Group Objective	Combined mean return	Combined Number	Surviving Funds mean return	Surviving Funds Number	Drop out %
1: Aggressive Growth	3.70	95	3.88	77	18.9
2: Growth/Small Companies	3.12	1052	3.22	904	14.1
3: Income/Growth-Income	2.49	540	2.59	434	19.6
4: Specialty	2.51	273	2.86	201	26.4
5: Foreign/World	2.04	490	2.18	429	12.4
6: Other	2.05	226	2.17	138	38.9

percentage of non-survivors. Moreover, the difference between the average annual returns for the 'specialty' category is 1.40%, which is much higher than the 0.64% for the aggregated sample of mutual funds.

6.4 What Determines Mutual Fund Survival?

In Section 6.2 we showed that the use of a survivorship free sample does not guarantee that standard estimators of mutual fund performance yield unbiased estimates. Moreover, as we briefly showed in Section 6.2 and what will be more extensively be discussed in Section 6.6, the use of a simple weight factor based on the ratio of the unconditional and conditional survival probability is sufficient to correct for survivorship bias in standard estimators. Consequently, in order to correct for survivorship bias we first of all have to determine the factors that affect mutual fund survival probabilities, which moreover, allows us to analyze the effects of survivorship on a variety of performance evaluation techniques.

In the previous section we noted that mutual funds that leave the sample have on average lower returns. Moreover, most of the disappearance occurs at a relatively young age, indicating that a bad record of returns in the first few years of its existence seriously decreases a fund's survival probabilities. It can also be noted that in particular months fund disappearance is much larger than can be expected on the basis of observed returns. To account for this, we include a common time effect in our specification.

[2] Note that a fund's investment objective is self-reported and can therefore easily lead to gaming to improve relative ex post return rankings (see Brown and Goetzmann [1997]).

Let y_{it} be an indicator variable that indicates whether or not fund i has an observed return in period t. Our first specification describes the probability of fund survival ($y_{it} = 1$) using a longitudinal probit model, such that a fund survives if an underlying latent variable, y^*_{it} is positive. That is,

$$\begin{aligned} y^*_{it} &= \alpha + \sum_{j=1}^{J} \gamma_{ij}(r_{i,t-j} - \theta) + \phi age_{i,t-1} + \lambda_t + \eta_{it} \\ y_{it} &= 1 \text{ if fund } i \text{ is observed in quarter } t \ (y^*_{it} > 0) \\ y_{it} &= 0 \text{ otherwise} \end{aligned} \tag{6.7}$$

where $r_{i,t-j}$ is the return of fund i in quarter $t-j$, θ is an unknown constant, $age_{i,t-1}$ is the time in years since fund inception, and λ_t denotes a time effect describing economy wide effects. The error term η_{it} is assumed to be standard normally distributed, independently over funds and periods. i.e. $\eta_{it} \sim IIN(0,1)$. The γ coefficients measure the impact of historical returns and, potentially, vary over funds and lags. To prevent that the model only applies to funds that have a return history of at least J quarters, and is thus conditional upon having survived these J quarters, we employ a flexible parametrization of the effects of lagged returns such that the model is conditional upon the observed return history only. In addition, to avoid dimensionality problems, we assume that the γ_{ij}'s can be described by a polynomial in j, multiplied by a factor that depends upon the number of lagged returns that is available. Let m_{it} denote the number of lagged quarterly returns that is available for fund i in quarter t, with a maximum of J. Then, we assume the following structure for the lagged quarterly returns coefficients[3]

$$\gamma_{ij} = (1 + \xi \ln [J + 1 - m_{it}]) \sum_{k=0}^{3} a_k j^k . I(j \leq m_{it}), \tag{6.8}$$

where $\sum_{k=0}^{3} a_k j^k$ is a polynomial of degree three, and $I(\,)$ is the indicator function that equals 1 if j is smaller than or equal to m_{it} and 0 otherwise. Note that for mutual funds with a return record of more than J quarters, the lagged quarterly returns coefficients can simply be described by $\sum_{k=0}^{3} a_k j^k$. The advantage of a polynomial lag structure is that we only have to estimate a restricted number of parameters, increasing precision of the estimates, and, moreover, a smooth pattern of the coefficients is automatically imposed. As it is implicitly assumed that further lags of the returns are irrelevant, we restrict the polynomial coefficients such that the hypothetical coefficient for lag $J + 1$ is zero. This gives an additional restriction that can be substituted in (6.7) and reduces the number of parameters describing variation in γ_{ij} to four. It should be

3 For notational simplicity, the fact that the γ_{ij} coefficients vary over time as a function of m_{it} (for a subsample of the funds) is not reflected in their indices.

noted, because of the presence of the time effects and the truncation of m_{it}, that both ξ and θ are only identified from information contained in funds that exist less than 12 quarters.

We estimate (6.7) with four ($J = 4$), eight ($J = 8$) and twelve ($J = 12$) lagged quarterly returns included, over the period 1989/01 through 1994/04. Table 6.4 reports the estimates for the lagged quarterly return parameters γ_1 through γ_J, the constant fund return parameter θ and the age parameter ϕ. The coefficient estimates for the time dummies can be found in Table 6.9 (Appendix 6.B). It appears that lagged quarterly returns, age of the fund and the

Table 6.4: **Estimation results.** The table presents estimation results for probit specification (6.7) with four ($J = 4$), eight ($J = 8$) and twelve ($J = 12$) lagged quarterly returns, a constant fund return θ, age of the fund (in years) and 24 time dummies as explanatory variables. We do not report the estimates for the polynomial coefficients, but only report the implied estimates for the lagged quarterly returns under the condition that a fund has more than J quarterly returns available. Note that for funds with less than J historical returns available, the coefficients for the lagged return should be inflated by a factor (see main text). The total number of observations is 36311.

	$J=4$		$J=8$		$J=12$	
	estimate	std. err	estimate	std. err	estimate	std. err
α	3.208	0.231	3.416	0.224	3.257	0.213
r_{t-1}	0.013	0.003	0.014	0.003	0.014	0.003
r_{t-2}	0.016	0.003	0.012	0.002	0.014	0.002
r_{t-3}	0.019	0.003	0.013	0.002	0.015	0.002
r_{t-4}	0.015	0.003	0.014	0.002	0.015	0.002
r_{t-5}	.		0.016	0.002	0.015	0.002
r_{t-6}	.		0.016	0.002	0.014	0.002
r_{t-7}	.		0.014	0.002	0.013	0.002
r_{t-8}	.		0.009	0.002	0.012	0.002
r_{t-9}	.		.		0.010	0.002
r_{t-10}	.		.		0.008	0.002
r_{t-11}	.		.		0.006	0.002
r_{t-12}	.		.		0.003	0.001
θ	11.394	2.229	8.390	1.215	6.878	0.795
ξ	-0.603	0.309	0.087	0.107	-0.066	0.064
age_{t-1}	0.016	0.004	0.019	0.004	0.025	0.005

aggregate time effect have a significant effect on fund disappearance (at the 5% level). Low returns for a number of consecutive quarters increase the probability of leaving the sample. The positive coefficient for age indicates that, *ceteris paribus*, the older the fund, the more likely it is to survive. It is also clear from the results that the probability of fund disappearance varies significantly over the quarters, even if returns have not changed. Note that the time effects capture all fund-invariant variables, including, for example, the return on the market portfolio and the term structure of interest rates. Most strikingly, during the third quarters of 1993 and 1994, fund disappearance has been much more likely than in other quarters.

The estimated values for ξ vary a lot between the specifications and are not significantly different from zero. This indicates that the absolute weights of recent historical returns do not increase for funds with a short history. Put differently, the returns in, say, the last four quarters are equally important irrespective of whether the fund has a history of just these four or more than twelve quarterly returns. The coefficient θ serves the purpose of adjusting the mean of the probit function when m_{it} changes, such that the number of returns included does not give a spurious effect on the survival probabilities through their nonzero means.

The three specifications in Table 6.4 are tested against each other and against more general alternatives. Note that the number of parameters in each of the three models is the same. To test whether the inclusion of additional lags would improve the models, we applied variable addition test to the three specifications. These tests are Lagrange Multiplier tests for the null hypotheses that the coefficients for one or more additional lags, added unrestrictedly to the model, are zero. More details about this and subsequent tests are provided in Appendix 6.A.

Panel A of Table 6.5 presents the outcome of the tests for the inclusion of additional lagged returns. Clearly, the specifications with $J = 4$ and $J = 8$ are overly restrictive and have to be rejected against alternatives with additional lags. For the model with three years of quarterly returns ($J = 12$) it cannot be rejected that further lags have zero coefficients. As it is well known that violation of the assumption of homoskedastic error terms typically leads to inconsistency of the maximum likelihood estimators in limited dependent variable models (see, e.g. Amemiya [1986 p. 268 ff.]), we also test the specification with $J = 12$ against heteroskedastic alternatives, the error variance being functions of lagged returns, fund age or both. The results of the Lagrange Multiplier tests, presented in Table 6.5, do not cause any doubt on the validity of the homoskedasticity assumption. Another crucial assumption is that of normality, which we tested against the more general Pearson family of distributions, as described by Newey [1985]. Somewhat surprisingly given the large number of observations, we are not able to reject normality either. Finally, we tested the inclusion of nonlinear functions of age and once more do not reject the model.

Let us now look at our preferred specification with $J = 12$ in more detail. For funds with a return history of less than 12 months, the coefficients in Table 6.4 are not appropriate and should be adjusted with the estimated factor $\xi \ln(J+1-m_{it})$ and set to zero for the unavailable returns. The results of this exercise are presented in Table 6.10 in Appendix 6.B. Using the estimates for the panel data probit model with 12 quarterly returns included, we can compute the probability that a fund will disappear in the next quarter given the past record of returns and the age of the fund. In Figure 6.1 we show the probability of disappearance for funds with different ages, where the past record of returns varies from -5% to $+5\%$ for each of the last four quarters and the quarterly returns for the quarters $t-5$ through $t-12$ are fixed at 3.00%, corresponding to the average quarterly return over the period 1989-1994. The probabilities are

Table 6.5: **Results of misspecification tests.** The table reports for the specifications (6.7) and (6.9) outcomes of LM tests for missing impacts of past performance of the fund, its style and the specification of age. All test statistics have an asymptotic null distribution that is Chi-squared with degrees of freedom given by DF. A * indicates rejection at the 5% level. See Appendix A for details

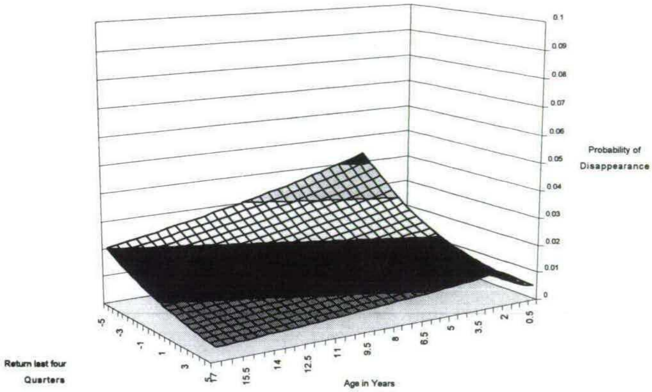
Panel A: Variable Addition Test					
Specification		Additional variable(s)	DF	LM-statistic	p-value
(6.7)	$J = 4$	r_{t-5}	1	7.43*	0.006
		$r_{t-5}...r_{t-8}$	4	45.99*	0.000
(6.7)	$J = 8$	r_{t-9}	1	7.91*	0.005
		$r_{t-9}...r_{t-12}$	4	12.89*	0.012
(6.7)	$J = 12$	r_{t-13}	1	1.77	0.183
		r_{t-13}, r_{t-14}	2	2.54	0.280
		$style\ dummies\ g_1...g_5$	5	57.61*	0.000
		age_{t-1}^2	1	2.28	0.131
		$\sqrt{age_{t-1}}$	1	3.54	0.060
(6.9)	$J = 12$	r_{t-13}	1	1.32	0.251
		r_{t-13}, r_{t-14}	2	2.01	0.366
		age_{t-1}^2	1	2.64	0.104
		$\sqrt{age_{t-1}}$	1	3.31	0.069
Panel B: Heteroskedasticity Test					
Specification		Variable(s)	DF	LM-statistic	p-value
(6.7)	$J = 12$	$r_{t-1}...r_{t-12}$	3	0.13	0.988
		$r_{t-1}...r_{t-12}$, age_{t-1}	4	0.21	0.995
(6.9)	$J = 12$	$r_{t-1}...r_{t-12}$, age_{t-1}	4	0.13	0.998
		$r_{t-1}...r_{t-12}$, $age_{t-1}, g_1..g_5$	9	0.43	0.999
Panel C: Normality Test					
Specification			DF	LM-statistic	p-value
(6.7)	$J = 12$		2	4.16	0.125
(6.9)	$J = 12$		2	2.82	0.244

averaged over the 24 different quarters. Alternatively, we could have fixed the time effect to its average over the quarters.

It appears that, for instance, a 3-year old fund with a return of -5% for each of the last four quarters has a probability of almost 4% to disappear in the next quarter, while a 16-year old fund with a comparable return record only has a probability of almost 2% to disappear. On the other hand, the probability of attrition drops below 1% if a fund of age sixteen had a return of +5% for the last four quarters.

The signs we find for the estimated coefficients in the probit specifications are in accordance with the results of Brown and Goetzmann [1995]. Our specification can be interpreted as a reduced form specification of their model, that also includes the size of the fund and the expense ratio as explanatory variables. While we do not observe the size of a fund during the entire sample period, it has been well documented (see, e.g., Rockinger [1995] and Sirri and Tufano [1997]) that (relative) historical returns are key determinants of capital flows to mutual funds. In contrast to our reduced form model, Brown and Goetzmann [1995] did not include time effects nor test for their presence. It is important to allow for fixed time effects to incorporate

Figure 6.1: **Non-Survival Probabilities.** The figure shows the probability of disappearance for different years since fund inception (Age). The return for each of the last four quarters, i.e. $r_{t-1}...r_{t-4}$, varies between -5% to +5%, while the returns over the quarters $r_{t-5}...\ r_{t-12}$ is fixed at 3% per quarter.



common aggregate shocks that affect the survival of all funds, such as bad returns on the stock market as a whole. Omitting the time effects, which may be correlated with the regressors, yields inconsistent parameter estimates (Baltagi [1995, p. 178 ff.]) and inappropriate bias corrections.

In order to examine whether mutual funds with different investment objectives have different probabilities of leaving the sample, we tested whether the inclusion of investment objective dummies significantly improves the model. Given that this test strongly rejects (see Table 6.5), we extended the survival model in (6.7) to include dummies for each investment objective. This leads to:

$$\begin{aligned} y_{it}^* &= \alpha + \sum_{j=1}^{J} \gamma_{ij}(r_{it-j} - \theta) + \phi age_{i,t-1} + \\ &\quad \delta_1 g_{1i} + \delta_2 g_{2i} + \delta_3 g_{3i} + \delta_4 g_{4i} + \delta_5 g_{5i} + \lambda_t + \eta_{it} \\ y_{it} &= 1 \text{ if fund } i \text{ is observed in quarter } t\ (y_{it}^* > 0) \\ y_{it} &= 0 \text{ otherwise} \end{aligned} \tag{6.9}$$

where g_{1i} through g_{5i} denote the investment objective dummies, corresponding to the classification in Table 6.3. As before, we assume that the structure for the lagged quarterly return coefficients γ_{ij} can be described by (6.8). Table 6.6 reports the estimates for specification (6.9) with twelve lagged quarterly returns included ($J = 12$), while the coefficient estimates for the time dummies can be found in Table 6.9 (Appendix 6.B). In Table 6.5, we also report the outcomes for the tests of the homoscedasticity and normality assumption in specification (6.9).

Table 6.6: **Estimation results Investment Categories dummies.** The table presents estimation results for probit specification (6.9) with twelve ($J = 12$) lagged quarterly returns, a dummy for the investment objective, a constant fund return θ, age of the fund (in years) and 24 time dummies as explanatory variables. We do not report the estimates for the polynomial coefficients, but only report the implied estimates for the lagged quarterly returns under the condition that a fund has more than 12 quarterly returns available. The total number of observations is 36311

	$J = 12$				
	estimate	std. err		estimate	std.err
α	3.161	0.232			
r_{t-1}	0.013	0.003	δ_1 : Growth	0.163	0.102
r_{t-2}	0.014	0.002	δ_2 : Income	0.090	0.105
r_{t-3}	0.015	0.002	δ_3 : Specialty	0.085	0.112
r_{t-4}	0.015	0.002	δ_4 : Foreign	0.310	0.112
r_{t-5}	0.015	0.002	δ_5 : Other	-0.287	0.073
r_{t-6}	0.015	0.002	θ	6.617	0.761
r_{t-7}	0.014	0.002	ξ	-0.063	0.068
r_{t-8}	0.013	0.002	$\phi : age_{t-1}$	0.024	0.005
r_{t-9}	0.011	0.002			
r_{t-10}	0.009	0.002			
r_{t-11}	0.006	0.002			
r_{t-12}	0.003	0.001			

From Table 6.6, it appears that U.S. based internationally investing mutual funds, i.e. investment objective 'foreign', have, *ceteris paribus*, the highest probability to survive. Moreover, the positive coefficients for the investment dummies in the majority of cases indicates that the mutual funds with investment objective 'aggressive growth' and investment objective summarized by the category 'other' have the highest probability to disappear. The estimated coefficients for the lagged returns, and age are similar to those for specification (6.7). While specification (6.9) describes survival probabilities conditional on a larger information set that includes investment objective, the significance of the investment dummies suggests that it can be expected that the error term η_{it} in (6.7) exhibits fund-specific serial correlation. We will therefore use specification (6.9) in the empirical analysis. For the Monte Carlo experiments in Sections 6.5 and 6.6, where we do not distinguish different investment styles, we use specification (6.7).

6.5 The Effects of Survivorship on Performance Measures

In Section 6.2 we examined in a stylized example the effect of non-random attrition on a simple performance measure like the average fund return using a survivorship free sample as well as a sample plagued by survivorship. Let us now look at a more realistic example, where interest lies in the estimation of fund alphas and their persistence. To do so, we perform a Monte Carlo simulation experiment. Following the set-up of Brown, Goetzmann, Ibbotson and Ross [1992], we generate quarterly returns from the following one factor model

$$r_{it} = r_{ft} + \beta_i(R_{mt} - r_{ft}) + u_{it} \tag{6.10}$$

where r_{ft} is the short term interest rate and $R_{mt} - r_{ft}$ is the quarterly excess return on the market portfolio. The idiosyncratic error term u_{it} is independent of the quarterly risk premium and is assumed to be normal with mean zero and variance σ_i^2, given by

$$\sigma_i^2 = k(1 - \beta_i)^2. \tag{6.11}$$

This relationship is a rough approximation to the relationship between non-systematic risk and β that is often observed in mutual funds data. We employ a set of parameter values in the return generating process that closely matches the first two sample moments of returns and beta. The quarterly excess return on a market portfolio is i.i.d. normal with mean 0.0215 and standard deviation 0.104, β is i.i.d. normal with mean 0.93 and standard deviation 0.37 and the value of k equals[4] 0.01997. Moreover, we assume that the short term interest rate can be described by an AR(1) model given by[5] [6]

$$r_{ft} = \mu + \rho(r_{ft-1} - \mu) + \epsilon_t, \quad \epsilon_t \; i.i.d. \; N(0, \sigma_\epsilon^2). \tag{6.12}$$

The simulation experiment proceeds as follows. We start with a 'random' number of funds such that an average 2% increase per quarter leads to 2500 mutual funds in the final quarter. This leads to a sample where the number of funds increases each quarter, while none of the funds drops out. Next, we apply the survival model of the previous section, i.e. equation (6.7) with twelve lagged returns ($J = 12$), to determine for each fund in each period the probability that it leaves the sample. This means that from the record of historical returns, the age of the fund and an aggregate time effect, a probability $\hat{p}_{it}$ of leaving the sample in the next quarter is determined. Then fund i leaves the sample in period t with a probability $\hat{p}_{it}$. Note that this assumes that η_{it} is independent of current returns, that is, the probability of survival only depends upon age and historical returns, and – conditional upon those – not upon current returns. Since the age of the fund, defined as the years since fund inception, is a significant factor in fund disappearance, we decided to draw a random age for the funds that already exist

[4] The value of k that we employ is based on the sample average of $\hat{\sigma}_i^2/(1 - \hat{\beta}_i)^2$.

[5] An OLS estimation of the short rate AR(1) process over the period 1976-1994 yields $r_{ft} = 0.140 + 0.922 * r_{ft-1} + \epsilon_t$, with $\hat{\sigma}_\epsilon = 0.003$.

[6] In the simulations, the average T-bill return over 1976-94 of 0.018 is used to start the process. Moreover, if a risk free rate smaller than zero happens to be generated, we set it equal to zero.

in the first quarter, closely corresponding to the observed age distribution in our sample of mutual funds, i.e. $age_1 \sim abs(N(0,16))$. The survivorship process used in our simulations is thus more complicated but also more realistic than the rules applied by Brown, Goetzmann, Ibbotson and Ross [1992] and Hendricks, Patel and Zeckhauser [1997], who simply remove, for instance, the worst performing 10% of the mutual funds in each period.

Note that the estimated time effects in probit specification (6.7) reflect aggregate shocks. We take the potential dependence between the time effects and observed aggregate variables in the model into account by running a number of regressions on variables such as the return on the S&P500 and the return on a three-month Treasury Bill over the period 1989-1994. Table 6.11 in Appendix 6.B presents the estimation results for a number of specifications. It appears that the time effect is significantly correlated with the risk-free return on Treasury bills. Accordingly, the random effect that is used in the simulations is drawn from a normal distribution with mean $\mu_{\lambda_t} = 2.29 + 0.332 * r_{ft}$ and variance equal to $\sigma_\lambda = 0.079$. Note that a high risk-free rate leads to, on average, higher time effects, but the effect on the survival rate might be balanced by higher (nominal) returns. The numbers presented below refer to averages or standard deviations over 500 replications. In order to prevent sensitivity to the starting conditions, we do not use the first 24 quarterly returns, i.e. fund returns are generated from quarter 1 onwards, while the survival process starts operating from quarter 13 and further.

We now construct four different samples. The first sample is the one without attrition and contains all funds up to the last period. We will refer to this hypothetical complete sample as "without" and we will only use it in a few cases. The second sample suffers from survivorship, as generated by our model, and contains only those funds that happened to survive until the end of the last quarter. We refer to this sample as "survivors". A third sample consists of the survivors sample completed with observations on those funds that left the sample before the last quarter. We refer to this sample as "combined". Most recent empirical studies (e.g. Brown and Goetzmann [1995], Carhart [1997a], Wermers [1997]) employ such survivorship free samples. A fourth and last sample named "non-survivors" is used for comparisons only and contains only the non-surviving funds.

First, Table 6.7 presents average quarterly returns over 36 quarters in the different samples. As expected, the mean return of the surviving funds substantially exceeds (i.e. 0.48% on an annual basis) the mean return for the combined or 'survivorship bias free' sample, at least if the parameter values in the simulation have been chosen to match the sample means. Furthermore, it appears that the non-surviving funds have, on average, a lower β than the surviving mutual funds.

Table 6.7: **Simulated average quarterly returns and betas** The table presents average quarterly returns and betas for 500 simulated samples of surviving funds, non-surviving mutual funds, the combined sample as well as the sample without attrition. Standard errors in parentheses. Averages are computed over 36 quarterly mean returns.

	Average return		Average β
Without	3.72	(0.02)	0.93
Survivors	3.80	(0.02)	0.93
Combined	3.68	(0.02)	0.92
Non-Survivors	2.42	(0.01)	0.90

Another important topic in performance analysis of mutual funds is the persistence in returns. Empirical studies by Brown and Goetzmann [1995], Malkiel [1995] and Carhart [1997a] examine whether 'winning' mutual funds, where winning is defined as exceeding the median fund return in a given period, are more likely to be winners in the next period. Studies of Brown, Goetzmann, Ibbotson and Ross [1992] and Hendricks, Patel and Zeckhauser [1997] show that survivorship bias induces spurious persistence patterns if there is cross-sectional variation in expected returns or risk. Instead of hypothesizing a certain survival process, we use the empirical survival model that matches the sample of U.S. equity funds, to redo the calculations of Hendricks, Patel and Zeckhauser [1997], who found a spurious J-shape pattern in performance persistence. As we generated fund returns such that any abnormal return is the result of (unpredictable) luck, any regularity found in performance measures is necessarily spurious and due to survivorship bias.

The performance of the funds is evaluated by estimating Jensen's α from the one-factor model in (6.10), over four three-year periods. We sort the funds on the basis of the estimated α's in each three-year period into eight groups. For each octile group, we calculate the average Jensen's α in the subsequent three-year period. Table 6.8 presents the average α for each group for the sample that only contains the surviving mutual funds, the sample that also contains the funds that ceased to exist before the final quarter, and the hypothetically complete sample, not affected by attrition.

It appears that the sample of surviving mutual funds exhibits a strong pattern of spurious persistence in performance. Furthermore, this is also the case for the combined sample, that includes funds that did not survive until the final quarter, although the pattern is somewhat weaker. Clearly, the fact that the data is survivorship free does not imply that a standard analysis is free of survivorship bias. Although the spurious persistence pattern is not exactly J-shaped, the simulation results more or less confirm the bias found by Hendricks, Patel and Zeckhauser [1997]. The argument for such a pattern is a risk argument. Funds in one of the extreme ranks are more likely to be 'high risk' funds and thus less likely to survive. Conditional on the fact

Table 6.8: **Simulated performance persistence.** The table presents the subsequent period performances for simulated samples of surviving funds, for samples that also contain the non-surviving funds until the quarter of disappearance and for samples that are not effected by survivorship. Standard errors in parentheses.

Initial Period Rank	Performance Subsequent Period Survivors		Combined		Without	
1	0.162	(0.007)	0.117	(0.006)	0.004	(0.005)
2	0.064	(0.005)	0.045	(0.004)	-0.001	(0.004)
3	0.035	(0.003)	0.024	(0.003)	0.001	(0.003)
4	0.018	(0.002)	0.014	(0.002)	0.001	(0.002)
5	0.024	(0.003)	0.017	(0.003)	-0.002	(0.002)
6	0.045	(0.004)	0.031	(0.004)	-0.002	(0.003)
7	0.064	(0.005)	0.042	(0.005)	-0.001	(0.004)
8	0.143	(0.008)	0.095	(0.008)	-0.000	(0.006)

that they did survive in the second subperiod, they will have made better returns than average. Compared to Hendricks, Patel and Zeckhauser, we find an additional upward bias in Jensen's α for the lower octiles. The reason for this is that our survival process is dynamic. Funds with a low rank realized relatively bad returns in the first twelve quarters. As this will decrease a fund's probability over the next twelve quarters to survive, these funds must have made up for these bad returns given that they have survived. Apparently, with our parameter values this effect is large enough to change the J-shape into a (more or less) U-shape. As expected, the sample that is not affected by survivorship does not exhibit a spurious persistence pattern.

6.6 Correcting for Survivorship Effects

Knowledge of the survival process is a key to correcting for the survivorship effects as discussed in the previous sections. In Section 6.4 survival of a fund was modelled as a function of historical returns, age and an aggregate time effect. We will show, in this section, how inferences can be corrected for survivorship effects if it can be assumed that fund survival in period t is independent of the return in period t, after conditioning upon lagged returns, fund age and time[7]. Technically, this imposes that η_{it} in (1) is independent of r_{it}, as was assumed in the previous section. The corrections, based upon work of Moffitt, Fitzgerald and Gottschalk [1997], are relatively simple to apply and involve the use of weights as shown in Section 6.2. As these weights depend upon fund returns, they are endogenous and their use has implications for consistency of the estimators used.

[7] Econometric approaches of sample selection and attrition problems based upon the work of Heckman [1979] and Hausman and Wise [1979], assume that the model of interest is conditional upon the same set of variables, which is inappropriate in this case.

In general, let R_i denote a vector of returns for fund i that is used in an empirical analysis, for example in constructing a contingency table. Let $Y_i = 1$ if fund i is used in the analysis and 0 otherwise. While Y_i is determined by the researcher, we shall assume that it is a function of y_{it}'s only. The distribution of returns for the funds used in the analysis, conditional on some observed characteristics X_i, is described by $f(R_i|X_i, Y_i = 1)$, where f is generic notation for a density/probability mass function. Because Y_i is not independent of the returns, this distribution differs from the one unconditional upon survival $f(R_i|X_i)$, which is what we are interested in. Here, X_i is chosen by the researcher and could be empty but could e.g. also reflect a fund's investment style. Let Z_i denote observable fund characteristics that affect the probability of survival. Then it follows, using standard conditioning arguments (see Moffitt et al. [1997]), that we can write

$$f(R_i, Z_i|X_i) = w_i f(R_i, Z_i|X_i, Y_i = 1), \tag{6.13}$$

where w_i is a weight factor given by

$$w_i = \frac{P\{Y_i = 1|X_i\}}{P\{Y_i = 1|R_i, X_i, Z_i\}}. \tag{6.14}$$

This weight equals the inverse of the normalized probability that fund i is kept in the sample for funds of type X_i. The left hand side of (6.13) provides the (un)conditional distribution of returns we are interested in. The right hand side is the conditional observable distribution of returns times a weight factor. If the weights w_i are known, any inference based upon the observed distribution of returns can directly be adjusted to reflect the unconditional distribution. For example, the expected returns of fund i satisfy

$$\begin{aligned} E[r_{it}] &= \\ \int R_i f(R_i, Z_i) dR_i dZ_i &= \int w_i R_i f(R_i, Z_i|Y_i = 1) dR_i dZ_i \\ &= E[w_i r_{it}|Y_i = 1], \end{aligned} \tag{6.15}$$

which implies that the average of $w_i r_{it}$ rather than the average of r_{it} provides an unbiased estimate of the fund's mean return if r_{it} is observed if $Y_i = 1$ only. Similarly, a fund's alpha can be estimated as

$$\tilde{\alpha}_i = w_i \hat{\alpha}_i, \tag{6.16}$$

where $\hat{\alpha}_i$ is the usual (uncorrected) ordinary least squares estimate.

Going back to our sample of U.S. equity funds, suppose we are interested in performance as measured by alpha over a period of 12 quarters. This implies that we can only use funds in the analysis that have observed returns for 12 consecutive periods s to $s + 11$, and we have that $Y_i = \prod_{t=s}^{s+11} y_{is}$. The probability that $Y_i = 1$ given the fund's returns R_i and characteristics Z_i, X_i is then described by our survival model, provided that X_i is included in the model (or can be assumed to have a zero coefficient), and provided that we assume that, conditional upon historical returns, Z_i and X_i, the probability of attrition in any given period does not depend upon (potentially unobserved) returns in that or future periods. In that case, we can write[8]

$$P\{Y_i = 1|R_i, Z_i, X_i) = \prod_{t=s}^{s+11} P\{y_{it} = 1|r_{i,t-1}, \ldots, age_i, style_i). \tag{6.17}$$

As we have estimated the latter probabilities from the sample of surviving and attrited funds, this provides estimates for the denominator in the weights w_i. The numerator in the weights reflects the probability of survival for a given X_i. When X_i is empty, and one is interested in returns for arbitrary funds, this can easily be estimated by the ratio of the number of funds that survived from period s to $s + 11$ and number of funds that was in the sample in period $s - 1$. If X_i denotes investment style, this computation has to be done for each style separately. Together, this provides estimated weights $\hat{w}_i$ that are consistent for $N \to \infty$. Consequently, we can estimate the alpha of an individual fund asymptotically unbiasedly[9] using (6.16) with $\hat{w}_i$ instead of w_i.

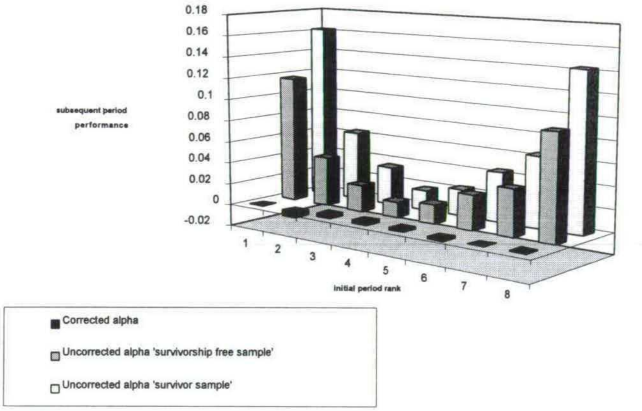
The approach above is generally applicable as long as it is clear what selection process a researcher is conditioning upon. In order to estimate unconditional expected returns in a given period t, for example, the conditioning is upon participation of each fund in that particular period and the weights simply reflect the probability that $y_{it} = 1$. Averaging over periods then, does not require additional corrections.

In order to illustrate the use of the proposed correction method, we applied it for possible biases in simulated samples generated in Section 6.5. Recall that in the performance persistence analysis we found a spurious U-shape pattern in the risk-adjusted returns of the formed octiles in the combined as well as survivors-only samples. To apply the correction approach to the contingency Table 6.8, we need to apply two corrections. First, to estimate initial period alphas and their ranking, we use funds that have observations over 12 consecutive quarters. Thus, we need to correct the OLS estimates using weights based upon these 12 periods. Second, for the alphas in the second subperiod, we only consider funds with a history of 24 consecutive quarters, so the weights will have to be based upon these 24 periods (even though the alphas reflect only 12 periods).

[8] This assumes that there is no autocorrelation in η_{it} in (3).

[9] Asymptotically unbiased means that the expected value of the estimator equals the true value if the number of funds N goes to infinity. The asymptotics underly consistent estimation of the survival process.

Figure 6.2: **Simulation results survivorship bias correction on alpha.** The figure shows the subsequent period performances for simulated samples of surviving funds (Uncorrected alpha 'survivor sample'), for samples that also contain the non-surviving funds (Uncorrected alpha 'survivorship free sample'), and for samples that contain the non-surviving funds but with performance correction for look-ahead bias (Corrected alpha).



In Figure 6.2 we present the average corrected Jensen's alpha in the subsequent three-year period for each octile group, where octile one represents the worst performing funds of the initial period. For comparison, the figure also contains the uncorrected results for the samples of surviving funds only and the samples that also contain the non-surviving funds until the quarter of disappearance, as given in Table 6.8.

It appears that the spurious persistence pattern that was present in the combined sample of funds has disappeared. Although not reported, the standard errors show that the average corrected Jensen's alphas of the octile groups are not significantly different from zero anymore. Furthermore, while the Monte Carlo results show how the spurious persistence pattern can be eliminated if there is no genuine persistence, the correction with weights can also be applied to obtain estimates of performance persistence that do not suffer from biases.

6.7 Empirical Results

A substantial number of empirical papers report persistence in the performance of mutual funds over one to three year horizons, see, e.g., Hendricks, Patel and Zeckhauser [1993], Gruber

[1996], Carhart [1997a] or Wermers [1997]. Mostly, these papers suggest that their results are free of survivorship bias and no attempts are made to correct for potential biases, apart from the inclusion of attrited funds' returns in the sample. In this section, we address the question of short-term predictability of mutual fund performance correcting for survivorship biases using the methodology discussed in the previous section.

As our sample of attrited funds goes back to only January 1989 we can estimate survival probabilities only over the period 1989/1-1994/4, and our survivorship bias free methodology is limited to equity funds over this period. Contrary to the simulation experiment, where a one-factor pricing model was adequate to price all assets, we cannot be sure about the model with respect to which risk-adjusted or abnormal returns should be defined. First, we shall apply a simple one-factor model, given by:

$$r_{i,t+1} - r_{f,t+1} = \alpha_i + \beta_i(r^m_{t+1} - r_{f,t+1}) + \varepsilon_{i,t+1}, \tag{6.18}$$

where $r_{i,t+1}$ is the return on mutual fund i in period $t+1$, r^m_{t+1} is the return on the market portfolio in period $t+1$ and $r_{f,t+1}$ is the return on a risk free asset. Second, we use Carhart's [1997a] four-factor model given by:

$$r_{i,t+1} - r_{f,t+1} = \alpha_i + \beta_{mi}(r^m_{t+1} - r_{f,t+1}) + \beta_{si} r^{smb}_{t+1} + \beta_{hi} r^{hml}_{t+1} + \beta_{pi} r^{pr1yr}_{t+1} + \varepsilon_{i,t+1}, \tag{6.19}$$

where r^{smb}_{t+1} is the difference between the returns on a portfolio of small stocks and one of big stocks, r^{hml}_{t+1} is the difference between the returns on a portfolio of high book-to-market and a portfolio of low book-to-market stocks and r^{pr1yr}_{t+1} is the difference between the return on a portfolio of stocks with the highest return over the previous year and a portfolio of stocks with the lowest return over the previous year[10]. We shall refer to α_i in (6.18) and (6.19) as the Jensen's alpha.

The question we try to answer is to what extent the ranking of a fund's alpha, over the subperiod 1989/1-1991/4, is informative about its alpha in the second subperiod 1992/1-1994/4. We do so by first estimating Jensen's alphas from (6.18) and (6.19) over the initial three-year period for all funds in the sample that survived these twelve quarters. These least squares estimates are biased because they are conditional upon survival. To correct for these biases, we employ the estimated survival probabilities as described by model (6.9) that also includes the investment style of a mutual fund. Using the technique of Section 6.6, we correct the estimated

[10] We are very grateful to Mark Carhart for providing the data with returns on the market index, SMB portfolio, HML portfolio and PR1YR portfolio.

α for survivorship bias using

$$\tilde{\alpha}_i = \frac{\hat{q}_s}{\prod_{s=1}^{s+11} \hat{p}_{is}} \hat{\alpha}_i, \tag{6.20}$$

where $\hat{p}_{is}$ is the estimated probability that fund i leaves the sample in period s, and $\hat{q}_s$ is the ratio of the number of funds in the same investment category as fund i that survived from period s to $s + 11$, and the number of funds in that category that was in the sample at $s - 1$. In the next step, we sort the funds into octiles on the basis of the corrected Jensen's α. For the subsequent three year period, i.e. 1992-1994, we estimate alphas again using only those funds that survived all 24 quarters. We correct the least squares estimates in the way indicated in (6.20), but now the correction weights are based on 24 quarters rather than 12. Finally, we compute the (unweighted) average within each octile.

The results are summarized in Figures 6.3 and 6.4. These figures present the average corrected Jensen's α in the subsequent three year period for each octile group of a number of different subsamples, where octile one represents the worst performing funds of the initial period. Figure 6.3 is based on the one-factor model given in (6.18), while Figure 6.4 represents the four-factor model given in (6.19). Both pictures show the persistence patterns for the funds with investment objectives 'growth', 'aggressive growth' and 'income', as well as that for the combined sample that combines these three investment objectives ('full sample'). Because neither factor model seems particularly adequate in explaining returns for funds in one of the remaining investment categories ('specialty', 'foreign' and 'other'), we excluded these categories from the figures.

For the one-factor model, the full sample of funds does not exhibit any positive persistence. Funds with a risk-adjusted return in the initial period that is below the median, have the highest Jensen's alpha in the subsequent period, and, on average, outperform the model by about 0.2% per quarter. On the other hand, the best performing funds of the initial period even have a negative average alpha in the evaluation period, corresponding to an underperformance of 0.6% per quarter. The result that we find for the one-factor model is in contrast with the strong evidence for a 'hot hand' phenomenon reported by Malkiel [1995]. Note that Malkiel used a survivorship free sample, but did not correct for the potential presence of survivorship bias. At a disaggregate level, 'growth' funds have a similar reverse pattern as the full sample of funds, while 'aggressive growth' funds show a negative U-shape pattern in the subsequent period. The 'income' funds do not exhibit a clear pattern of performance persistence, but it seems that the best performing funds of the initial period belong to the worst performing in the second period.

However, if we move away from the one-factor model and concentrate on Carhart's four-factor model, we find that for the full sample of funds, as well as its subsamples, the (reverse)

Figure 6.3: **Empirical results for the bias corrected one-factor performance persistence pattern.** The figure shows the subsequent period performance measured by a one-factor model for the full sample of funds and for the funds with as investment objective: growth, aggressive growth and income. All results are corrected for bias using the procedure outlined in the text.

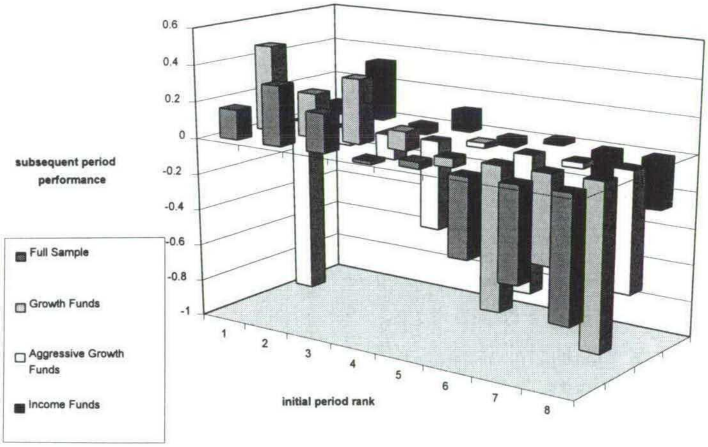
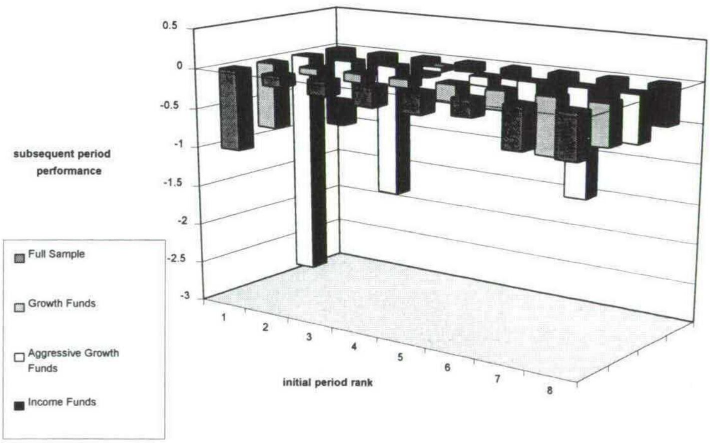


Figure 6.4: **Empirical results for the bias corrected four-factor performance persistence pattern.** The figure shows the subsequent period performance measured by Carhart's four-factor model, for the full sample of funds, and for the funds with as investment objective: growth, aggressive growth and income. All results are corrected for bias using the procedure outlined in the text.



persistence patterns have disappeared. There is no octile for any of the subsamples that significantly outperforms the model. Alternatively, this can be interpreted that the four factors in (6.19) account for the reverse persistence pattern of the one-factor model. It appears that the worst performing funds of the initial period are also the worst performing in the subsequent period, with an underperformance of 1% per quarter, implying persistence of bad performance for this group of funds. The pattern we find for the four-factor model is in accordance with the one reported by Carhart [1997a], indicating that our results do not support a hot hand phenomenon in mutual fund performance. However, in contrast to our findings, the best performing funds of Carhart's initial period have a slightly positive Jensen's alpha in the subsequent period. Moreover, the worst performing funds of Carhart's sample show less underperformance than the worst performing funds in the sample that we employ. There are two possible explanations for this difference in result. First of all, the difference in sample period, i.e. 1966-1993 vs 1989-1994. Second, recalling the spurious persistence pattern we found in the simulation experiment, we found that the worst performing funds had a higher persistence bias than the best performing mutual funds. Since Carhart's methodology is not free of this survivorship bias (look ahead bias), the difference in the two extreme octiles might be due to this effect.

6.8 Concluding Remarks

In the recent literature, the importance of survivorship bias in empirical studies in finance has been sufficiently acknowledged. Most studies emphasize the potential biases that can arise from analyzing data conditional upon survival, using more or less ad hoc theoretical models that determine survival. Attempts to correct for these biases are scarce and this chapter fills this gap.

We showed how inferences on mutual fund performance can be corrected for survivorship bias using a simple weighting strategy, based upon the estimated survival model. A Monte Carlo experiment showed that the spurious U-shape pattern that arises in estimating performance persistence using traditional techniques disappears with the correction that we propose. In addition, the approach was applied to U.S. equity funds using a one-factor and Carhart's four-factor model. Using the latter model, we do not find any evidence for 1989-1994 of the hypothesis that mutual funds that performed well in the past continue to perform well in the future.

In the chapter, we analyzed the potential effect of survivorship bias on various mutual fund performance measures, on the basis of an empirical model of survival, fitted to U.S. equity funds over the period 1989/1-1994/4. This required us to extend the study of Brown and Goetzmann [1995], by examining which factors are important in determining mutual fund survival probabilities. From an extensive analysis of various specifications, it appeared that a specification with twelve lagged quarterly returns, time since fund inception, aggregate time effects and dummies reflecting the investment style has to be preferred. The specification of the survival model was chosen in such a way that it is not conditional upon the existence of a three-year history of returns, so that it also models survival for 'young' funds.

In order to obtain insight in the size of survivorship effects in various performance evaluation measures, a number of Monte Carlo simulation experiments have been performed. By dropping funds from the sample based on the estimated survival probabilities, we analyzed the effect of survivorship on average returns and persistence in risk-adjusted returns, thus extending the analysis in Hendricks, Patel and Zeckhauser [1997]. Although the results are sensitive to the parameter values of the return generating process, we find that, as expected, average returns of samples of surviving funds only, are biased upward. Both the sample with surviving funds, as well as the sample that include returns of both survived and attrited funds, are affected by survivorship bias and generate a spurious persistence in performance. This is important, as it is generally believed and suggested that such survivorship free samples are free of survivorship bias. With the dynamic survival model that was estimated, a spurious U-shape pattern was found in the persistence of risk-adjusted returns, similar to but different from J-shape found in Hendricks, Patel and Zeckhauser [1997].

Appendix 6.A

Misspecification Tests in the Probit Model

In this appendix we briefly indicate how the different misspecification tests for the probit model have been computed. In particular, we consider Lagrange multiplier (or conditional moment) tests for omitted variables, heteroskedasticity and nonnormality. More details can be found in, e.g., Newey [1985] or Pagan and Vella [1989].

Variable addition tests

Let x_{it} denote the k-dimensional vector of explanatory variables in the probit model, including the time dummies. The log likelihood function for the probit model is given by

$$L(\beta|X,y) = \sum_{i,t} y_{it} \log \Phi(x_{it}'\beta) + \sum_{i,t} (1 - y_{it}) \log(1 - \Phi(x_{it}'\beta)), \tag{A.1}$$

so that the first order conditions can be written as

$$\sum_{i,t} \left[\frac{y_{it} - \Phi(x_{it}'\hat{\beta})}{\Phi(x_{it}'\hat{\beta})(1 - \Phi(x_{it}'\hat{\beta}))} \phi(x_{it}'\hat{\beta}) \right] x_{it} = \sum_{i,t} \hat{\varepsilon}_{it}^G x_{it} = 0, \tag{A.2}$$

where ϕ is the standard normal density function and Φ is the corresponding distribution function. The term in square brackets is referred to as the generalized residual (see Gourieroux et al. [1987]) and denoted $\hat{\varepsilon}_{it}^G$. The first order conditions can be interpreted to say as that each explanatory variable should be orthogonal to the generalized residual (over the whole sample).

If r additional variables z_{it} were to be included in the model, it would not change the current estimates if the current estimates already satisfy the additional first order conditions. This means that if

$$\sum_{i,t} \hat{\varepsilon}_{it}^G z_{it} = 0 \tag{A.3}$$

then including z_{it} in the model would not change the current estimates. To test whether the left hand side of (A.3) significantly differs from zero, we compute the Lagrange Multiplier test statistic as

$$\xi_{LM} = \iota' R (R'R)^{-1} R' \iota \tag{A.4}$$

where R is a matrix of individual gradients of the loglikelihood function, with typical row

$$(\hat{\varepsilon}_{it}^G x_{it}', \hat{\varepsilon}_{it}^G z_{it}'),$$

and ι is a vector of ones. It can be shown that under the null hypothesis that z_{it} does not enter the probit specification in (A.1), the Lagrange multiplier test ξ_{LM} is asymptotically χ^2 distributed with r degrees of freedom.

Testing for heteroskedasticity

Suppose that ε_{it} has a variance of

$$V[\varepsilon_{it}] = h_{it} = h(z'_{it}\gamma)$$

for some function $h > 0$ with $h(0) = 1$ (normalization condition), where z_{it} is of dimension r. Using this specification for the variance the loglikelihood changes to the following form

$$\begin{aligned} L(\beta, \gamma | X, y) &= \sum_{i,t} y_{it} \log \Phi \left(\frac{x'_{it}\beta}{\sqrt{h(z'_{it}\gamma)}} \right) + \\ &\quad \sum_{i,t} (1 - y_{it}) \log \left(1 - \Phi \left(\frac{x'_{it}\beta}{\sqrt{h(z'_{it}\gamma)}} \right) \right). \end{aligned} \tag{A.5}$$

Now the first order conditions for γ, evaluated under the null hypothesis $H_0 : \gamma = 0$ are

$$\sum_{i,t} \hat{\varepsilon}^G_{it}(x'_{it}\beta) z_{it} = 0.$$

Consequently, it is easy to test $H_0 : \gamma = 0$ using the Lagrange Multiplier test statistic given in (A.4) using a matrix R that has typical row

$$(\hat{\varepsilon}^G_{it} x'_{it}, \hat{\varepsilon}^G_{it}(x_{it}\hat{\beta}) z'_{it}).$$

Testing for non-normality

A test for normality can be derived by specifying an the alternative distribution function as $\Phi(x'\beta + \gamma_2(x'\beta)^2 + \gamma_3((x'\beta)^3)$ (compare Newey, 1985). The null hypothesis of normality corresponds to $\gamma_2 = \gamma_3 = 0$. This can be tested by using (A.4), where the matrix R now contains

$$(\hat{\varepsilon}^G_{it} x'_{it}, \hat{\varepsilon}^G_{it}(x'_{it}\beta)^2, \hat{\varepsilon}^G_{it}(x'_{it}\beta)^3).$$

Appendix 6.B

Additional Tables

Table 6.9: **Estimates time dummy coefficients.** The table reports the estimates for the time dummies for a probit specification with four ($J = 4$), eight ($J = 8$) and twelve ($J = 12$) lagged quarterly returns, and age of the fund (in years) as explanatory variables. The column $J = 12^*$ contains the estimates for the time dummies with as additional explanatory variable a dummy for the investment objective.

	$J = 4$		$J = 8$		$J = 12$		$J = 12^*$	
λ	estimate	std. err	estimate	std. err	estimate	std. err	estimate	std.err
89/02	-0.067	0.259	-0.006	0.261	-0.115	0.260	-0.091	0.263
89/03	-0.289	0.240	-0.266	0.242	-0.351	0.240	-0.341	0.240
89/04	-0.575	0.230	-0.555	0.233	-0.570	0.232	-0.544	0.232
90/01	-0.118	0.275	-0.227	0.275	-0.075	0.279	-0.054	0.280
90/02	-0.182	0.248	-0.306	0.252	-0.176	0.257	-0.165	0.258
90/03	-0.544	0.218	-0.771	0.219	-0.652	0.224	-0.627	0.225
90/04	-0.120	0.231	-0.450	0.230	-0.347	0.234	-0.336	0.235
91/01	0.070	0.243	-0.238	0.241	-0.147	0.242	-0.160	0.243
91/02	-0.572	0.220	-0.796	0.215	-0.744	0.215	-0.725	0.217
91/03	-0.626	0.226	-0.719	0.212	-0.751	0.213	-0.746	0.214
91/04	-0.449	0.233	-0.287	0.236	-0.355	0.235	-0.366	0.235
92/01	-0.636	0.220	-0.576	0.225	-0.598	0.223	-0.616	0.223
92/02	-0.522	0.219	-0.631	0.223	-0.584	0.221	-0.606	0.222
92/03	-0.483	0.218	-0.638	0.221	-0.539	0.223	-0.561	0.224
92/04	-0.510	0.212	-0.818	0.213	-0.634	0.218	-0.654	0.219
93/01	-0.287	0.222	-0.598	0.220	-0.476	0.224	-0.509	0.224
93/02	-0.193	0.233	-0.372	0.230	-0.345	0.233	-0.369	0.233
93/03	-1.286	0.202	-1.410	0.201	-1.407	0.203	-1.442	0.203
93/04	-0.648	0.211	-0.736	0.212	-0.772	0.213	-0.804	0.214
94/01	-0.476	0.215	-0.619	0.217	-0.631	0.217	-0.656	0.218
94/02	-0.533	0.210	-0.715	0.212	-0.684	0.212	-0.712	0.213
94/03	-0.980	0.201	-1.200	0.202	-1.145	0.204	-1.182	0.204
94/04	-0.311	0.212	-0.592	0.210	-0.535	0.213	-0.560	0.213

Table 6.10: **Quarterly return coefficients (* 100).** The table presents the implied estimates in the probit specification for survival probabilities for the quarterly return coefficients multiplied by 100 for funds with less than $J = 12$ returns available (m_{it})

pm.\ml	1	2	3	4	5	6	7	8	9	10	11	12
γ_1	1.14	1.15	1.16	1.17	1.18	1.19	1.20	1.22	1.24	1.27	1.30	1.36
γ_2	0	1.21	1.22	1.23	1.24	1.25	1.27	1.29	1.31	1.33	1.37	1.44
γ_3	0	0	1.25	1.26	1.28	1.29	1.30	1.32	1.34	1.37	1.41	1.48
γ_4	0	0	0	1.27	1.28	1.29	1.31	1.33	1.35	1.38	1.42	1.48
γ_5	0	0	0	0	1.26	1.27	1.28	1.30	1.32	1.35	1.39	1.46
γ_6	0	0	0	0	0	1.22	1.23	1.25	1.27	1.29	1.33	1.39
γ_7	0	0	0	0	0	0	1.14	1.16	1.18	1.20	1.24	1.30
γ_8	0	0	0	0	0	0	0	1.04	1.06	1.08	1.11	1.17
γ_9	0	0	0	0	0	0	0	0	0.91	0.93	0.96	1.00
γ_{10}	0	0	0	0	0	0	0	0	0	0.75	0.77	0.80
γ_{11}	0	0	0	0	0	0	0	0	0	0	0.54	0.57
γ_{12}	0	0	0	0	0	0	0	0	0	0	0	0.30

Table 6.11: **Dependence time effect and aggregate variables.** The table presents the estimation results for a regression of the estimated time dummies λ_1 through λ_{24} on a number of fund invariant variables. The variable c denotes a constant.

Independent Variables	Estimate	Std. error	adjR^2
c	2.296	0.155	
Treasury Bill	0.333	0.110	0.26
c	2.677	0.074	
S&P 500	0.017	0.015	0.32
c	2.289	0.154	
Treasury Bill	0.308	0.111	
S&P 500	0.012	0.010	0.28

Chapter 7
Summary

The main topic of this thesis was testing the validity of some of the motivations that have been given in the literature for investing in mutual funds. One of the motivations that gets much attention in the literature is the claim that the managers of these mutual funds have special abilities in selecting stocks, which makes the fund an interesting investment product with return characteristics that were not attainable by individual investors themselves. In order to test whether investors can extend the mean-variance efficient investment set of their current portfolio by taking a position in a mutual fund, the performances of mutual funds are evaluated. Chapter 2 introduced the main concepts of this thesis. It was shown that both the hypothesis that there is only one value of the risk aversion coefficient for which the investor cannot extend his investment set by taking a position in a mutual fund and the hypothesis that this is the case independent of the investor's risk attitude can simply be tested in a regression framework. Alternatively, these tests can be interpreted as testing for outperformance of the mutual funds with respect to a number of benchmark assets, where it is implicitly assumed that the investor already holds an efficient combination of the benchmark assets corresponding to a pricing model. Under specific assumptions, a similar regression framework can be used to test for outperformance conditional on economic circumstances.

Tests of the hypothesis that funds that performed well in the past continue to perform well, i.e. the hypothesis of persistence in performance, are sensitive to the choice of the factor mimicking portfolios included in the performance evaluation model. Exposure to common factors, such as for instance a momentum factor, can explain much of the observed persistence in mutual fund returns. Alternatively it can be stated that investors that are not exposed to this factor can extend the investment set by taking a position in the funds that actively follow a momentum strategy. Finally, it is shown that survivorship effects are present in evaluating mutual fund performance and its persistence, using traditional techniques.

Chapter 3 analyzed the performance of internationally investing U.S.- based mutual funds, correcting for market frictions such as short sales constraints and transaction costs. We show that if market frictions are absent, a risk-averse mean-variance optimizing U.S. investor can extend his investment set with an internationally investing mutual fund, given that the investor initially holds a widely diversified international portfolio with predetermined country weight allocation corresponding to the market capitalization of the individual countries. For some of the funds this finding is robust for incorporating short sales restrictions and transaction costs.

However, for investors that currently invest efficiently in the American, European and Japanese stock indices, an extension of the investment set by a mutual fund is not present anymore after imposing short sales restrictions on the benchmark assets as well as on the mutual funds. Finally, we show that market frictions can easily be incorporated in tests for conditional outperformance, where the conditioning is upon a set of information variables assumed to reflect the state of the economy. It is shown that some of the internationally investing mutual funds still provide an extension of the investment set for investors with a predetermined country weight allocation, although the economic circumstances can have a significant effect on this conclusion.

A simple way to evaluate performance is a comparison of a realized mutual fund return with a return on a benchmark asset that reflects the investment style of the mutual fund under consideration. Often, an index that reflects the self-reported style of the fund is used in this so-called relative performance evaluation. Chapter 4 showed that return-based style analysis can be used to improve relative performance evaluation. First of all, style analysis can be used to objectively determine the fund's investment style, and strongly related, style analysis can be used as an instrument in relative performance evaluation. A relative as well as risk-adjusted performance evaluation of a sample of Dutch mutual funds shows that taking into account the cash position the mutual funds have, seriously affects the relative performance of mutual funds. It is shown that under the assumptions that the fund manager's ability is independent of the benchmark asset return and that the exposure to the benchmark assets is equal to one, relative performance evaluation is an appropriate method to evaluate mutual funds. Most strikingly, funds that mainly invest in the Dutch stock market provide an extension of the mean-variance efficient set for Dutch investors that currently hold an internationally diversified portfolio, even after imposing short sell restrictions.

Chapter 5 analyzed small sample biases that arise in some estimation methods used for testing for predictability of mutual fund returns. A Monte Carlo simulation experiment shows that these biases are not at all negligible. The sign of the bias does not generate a spurious 'hot hands' phenomenon. Two approaches are suggested to eliminate the bias. The first one requires the estimation of expected returns over future observations, what implies that the most recent observations are not used in the estimation of the short-run persistence coefficients. The second approach is based on instrumental variables estimation of the model in first differences. Although this method yields consistent estimates, the Monte Carlo experiment shows that it leads to rather inefficient estimates. An empirical analysis suggests that a small persistence effect is present in mutual funds that mainly invest in U.S. growth stocks. In particular, the results suggest that selecting in each quarter the funds with high returns, relative to other funds, over the last four quarters, can significantly increase the expected return on a portfolio of mutual funds.

Finally, Chapter 6 analyzed whether mutual fund performance evaluation studies and studies in performance persistence are affected by survivorship effects. Using a so-called survivorship free sample, it is shown that nevertheless survivorship biases arise in evaluating mutual fund performances and its persistence using standard techniques. In order to correct for survivorship bias, knowledge of the survival process is essential. Therefore, a process is modelled that determines mutual fund survival. It is shown that the past record of returns, the age of the fund and aggregate macro-economic shocks significantly affect mutual fund survival probabilities. A Monte Carlo experiment shows that the survival process leads to a spurious pattern of performance persistence. It is shown that the survival model can be used to construct weights that can be applied to correct for survivorship biases in performance measurement methods. The correction method is used to estimate persistence in performance of U.S.-based growth, aggressive growth and income funds. Using a four-factor model with factor mimicking portfolios for a size effect, a book-to-market effect, a momentum effect and a market portfolio, it is shown that there is no evidence for the claim that mutual funds that performed well in the past continue to perform well.

References

Admati, A.R., S. Bhattacharya, P. Pfleiderer and S.A. Ross, 1986, On Timing and Selectivity, *Journal of Finance*, 41, 715-732.

Amemiya, T., 1985, *Advanced Econometrics*, Cambridge: Harvard University Press.

Anderson, T.W. and C. Hsiao, 1982, Formulation and Estimation of Dynamic Models using Panel Data, *Journal of Econometrics,* 18, 47-82.

Baltagi, B.H., 1995, *Econometric Analysis of Panel Data*, Wiley.

Bekaert, G. and M.S. Urias, 1996, Diversification, Integration, and Emerging Market Closed-End Funds, *Journal of Finance*, 51, 835-870.

Blake, C.R., E.J. Elton and M.J. Gruber, 1993, The Performance of Bond Mutual Funds, *Journal of Business*, 66, 371-403.

Brown, S.J., W.N. Goetzmann, R.G. Ibbotson and S.A. Ross, 1992, Survivorship Bias in Performance Studies, *Review of Financial Studies*, 5, 553-580.

Brown, S.J. and W.N. Goetzmann, 1995, Performance Persistence, *Journal of Finance*, 50, 679-698.

Brown, S.J., W.N. Goetzmann and S.A. Ross, 1995, Survival, *Journal of Finance*, 50, 853-874.

Brown, S.J. and W.N. Goetzmann, 1997, Mutual Fund Styles, *Journal of Financial Economics*, 43, 373-399.

Carhart, M.M., 1997a, On Persistence in Mutual Fund Performance, *Journal of Finance*, 52, 57-82.

Carhart, M.M., 1997b, Mutual Fund Survivorship, Working Paper, Marshall School of Business, U.S.C.

Chan, L.K., N. Jegadeesh and J. Lakonishok, 1996, Momentum Strategies, *Journal of Finance*, 51, 1681-1714.

Chen, Z. and P.J. Knez, 1996, Portfolio Performance Measurement: Theory and Applications, *Review of Financial Studies*, 9, 511-556.

Christopherson, J.A., W.E. Ferson and D.A. Glassman, 1995, Conditioning manager Alphas on economic Information: Another Look at the persistence of Performance, working paper.

Cochrane, J.H., 1997, Asset Pricing, manuscript, Graduate School of Business, University of Chicago.

Cumby, R.E. and J.C. Glen, 1990, Evaluating the Performance of International Mutual Funds, *Journal of Finance*, 45, 497-521.

Daniel, K., M. Grinblatt, S. Titman and R. Wermers, 1997, Measuring Mutual Fund Performance with Characteristic-Based Benchmarks, *Journal of Finance*, 52, 1035-1058.

DeRoon, F.A., 1997, Essays on Testing for Spanning and on Modelling Futures Risk Premia, *Tilburg University Thesis*, 33.

DeRoon, F.A., Th.E. Nijman and B.J.M. Werker, 1996, Testing for Spanning with Futures Contracts and Nontraded Assets: A General Approach, *CentER Discussion Paper no. 9683*, Tilburg University.

DeRoon, F.A., Th.E. Nijman and B.J.M. Werker, 1998, Testing for Mean-Variance Spanning with Short Sales Constraints and Transaction Costs: The Case of Emerging Markets, *CentER Discussion Paper no. 9807*, Tilburg University.

DeSantis, G., 1994, Asset Pricing and Portfolio Diversification: Evidence from Emerging Financial Markets, in Howell, M, (ed.): *Investing in Emerging Markets*, Euromoney Books, Londen.

Elton E.J., M.J. Gruber and C.R. Blake, 1995, Survivorship Bias and Mutual Fund Performance, *The Review of Financial Studies*, 9, 1097-1120.

Elton, E., M.J. Gruber and C.R. Blake, 1996, The Persistence of Risk-Adjusted Mutual Fund Performance, *Journal of Business*, 69, 133-157.

Fama, E.F., 1996, Multifactor Portfolio Efficiency and Multifactor Asset Pricing, *Journal of Financial and Quantitative Analysis*, 31, 441-465.

Fama, E. and K.R. French, 1993, Common Risk Factors in the Returns on Stocks and Bonds, *Journal of Financial Economics*, 33, 3-53.

Fama, E. and K.R. French, 1996, Multifactor Explanations of Asset Pricing Anomalies, *Journal of Finance*, 51, 55-84.

Fama, E. and J. Macbeth, 1973, Risk, Return, and Equilibrium: Empirical Test, *Journal of Political Economy*, 81, 607-636.

Farnsworth H., W. Ferson, D. Jackson, S. Todd, and B. Yomtov, 1995, Conditional Performance Evaluation, working paper.

Ferson, W.E. and C.R. Harvey, 1993, The Risk and Predictability of International Equity Market Returns, *Review of Financial Studies*, 6, 527-566.

Ferson W.E. and R.W. Schadt, 1996, Measuring Fund Strategy and Performance in Changing Economic Conditions, *Journal of Finance*, 51, 425-461.

Fung, W. and D.A. Hsieh, 1997, Empirical Characteristics of Dynamic Trading Strategies: The Case of Hedge Funds, *Review of Financial Studies*, 10, 275-302.

Goetzmann, W.N. and R.G. Ibbotson, 1994, Do Winners Repeat? Patterns in Mutual Fund Return Behavior, *Journal of Portfolio Management*, 20, 9-18.

Gourieroux C., A. Monfort, E. Renault, A. Trognon, 1987, Generalized Residuals, *Journal of Econometrics*, 34, 5-32.

Grinblatt M., and S. Titman, 1989a, Mutual Fund Performance: An Analysis of quarterly Portfolio Holdings, *Journal of Business*, 62, 393-416.

Grinblatt M., and S. Titman, 1989b, Portfolio Performance Evaluation: Old Issues and New Insights, *Review of Financial Studies*, 2, 393-421.

Grinblatt M., and S. Titman, 1992, The Persistence of Mutual Fund Performance, *Journal of Finance*, 47, 1977-1984.

Grinblatt M., and S. Titman, 1993, Performance Measurement without Benchmarks: An Examination of Mutual Fund Returns, *Journal of Business*, 66, 47-68.

Gruber, M.J., 1996, Another Puzzle: The Growth in Actively Managed Mutual Funds, *Journal of Finance*, 51, 783-811.

Hansen, L.P., J. Heaton and E.H.J. Luttmer, 1995, Econometric Evaluation of asset Pricing Models, *Review of Financial Studies*, 8, 237-274.

Hansen, L.P., and R. Jagannathan, 1991, Implications of Security Market Data for Models of Dynamic Economies, *Journal of Political Economy*, 99, 225-262.

Hausman, J.A. and D. Wise, 1979, Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment, *Econometrica*, 47, 455-473.

Heckman, J., 1976, The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models, *The Annals of Economic and Social Measurement*, 5, 475-492.

Heckman, J., 1979, Sample Selection Bias as a Specification Error, *Econometrica*, 47, 153-161.

Hendricks, D., J. Patel and R. Zeckhauser, 1993, Hot Hands in Mutual Funds: Short-Run Persistence of Relative Performance, 1974-1988, *Journal of Finance*, 48, 93-130.

Hendricks, D., J. Patel and R. Zeckhauser, 1997, The *J*-shape of Performance Persistence Given Survivorship Bias, *The Review of Economics and Statistics*, 79, 161-166.

Henriksson, R.D., 1984, Market Timing and Mutual Fund Performance: An Empirical Investigation, *Journal of Business*, 57, 73-96.

Hsiao, C., 1985, *Analysis of Panel Data*, Cambridge University Press.

Huberman, G. and S. Kandel, 1987, Mean-Variance Spanning, *Journal of Finance*, 42, 873-888.

Ibbotson, R.G. and R. Sinquefeld, 1990, *Stocks Bonds and Inflation: 1990 Yearbook*, Ibbotson Associates, Chicago.

Jegadeesh, N., 1990, Evidence of Predictable Behavior of Security Returns, *Journal of Finance*, 45, 881-898.

Jegadeesh, N. and S. Titman, 1993, Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency, *Journal of Finance*, 48, 65-91.

Jensen, M.C., 1968, The Performance of Mutual Funds in the period 1945-1964, *Journal of Finance*, 23, 389-416.

Jobson, J.D. and B. Korkie, 1989, A Performance Interpretation of Multivariate Tests of Intersection, Spanning and Asset Pricing, *Journal of Financial and Quantitative Analysis*, 24, 285-304.

Keim, D.B. and R.F. Stambaugh, 1986, Predicting Returns in the Bond and Stock Markets, *Journal of Financial Economics*, 17, 357-390.

Kodde, D.A. and F.C. Palm, 1986, Wald Criteria for Jointly Testing Equality and Inequality Restrictions, *Econometrica*, 54, 1243-1248.

Korajczyk, R.A. and C.J. Viallet, 1989, An Empirical Investigation of International Asset Pricing, *Review of Financial Studies*, 2, 553-585.

Luttmer, E.G.J., 1996, Asset Pricing in Economies with Frictions, *Econometrica*, 64, 1429-1467.

Malkiel, B.G., 1995, Returns from Investing in Equity Mutual Funds 1971 to 1991, *Journal of Finance*, 50, 549-572.

Markowitz, H.M., 1991, *Portfolio Selection*, Basil Blackwell, Cambridge, Massachusetts.

Moffitt, R., J. Fitzgerald and P. Gottschalk, 1997, Sample Attrition in Panel Data: The Role of Selection on Observables, paper presented at the 7th International Conference on Panel Data, Paris, 19-20 June, 1997.

Newey, W.K., 1985, Maximum Likelihood Specification Testing and Conditional Moment Tests, *Econometrica*, 53, 1047-1070.

Pagan, A. and F. Vella, 1989, Diagnostic Tests for Models Based on Individual Data: A Survey, *Journal of Applied Econometrics*, 4, S29-S59.

Patel, J., R. Zeckhauser, and D. Hendricks, 1992, Investment Flows and Performance: Evidence from Mutual Funds, Cross-Border Investments, and New Issues, in R. Sato, R.

Levich and R. Ramachandran, eds., *Japan and International Financial Markets: Analytical and Empirical Perspectives*, Cambridge University Press.

Rockinger, M., 1995, Determinants of Capital Flows to Mutual Funds, Working Paper, HEC, France.

Sirri, E.R. and P. Tufano, 1997, Costly Search and Mutual Fund Flows, forthcoming *Journal of Finance.*

Shanken, J., 1990, Intertemporal Asset Pricing: An Empirical Investigation, *Journal of Econometrics*, 45, 99-120.

Sharpe, W.F., 1965, Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk, *Journal of Finance*, 19, 425-442.

Sharpe, W.F., 1992, Asset Allocation: Management Style and Performance Measurement, *Journal of Portfolio Management*, Winter, 7-19.

Shukla, R.K. and G.B. van Inwegen, 1995, Do Locals Perform Better than Foreigners?: An Analysis of UK and US Mutual Fund Managers, *Journal of Economics and Business*, 47, 241-254.

Treynor, J. and K. Mazuy, 1966, Can Mutual Funds Outguess the Market?, *Harvard Business Review*, 44, 131-136.

Wermers, R., 1997, Momentum Investment Strategies of Mutual Funds, Performance Persistence, and Survivorship Bias, working paper

Samenvatting (Summary in Dutch)

Veel beleggers kiezen ervoor te beleggen via beleggingsfondsen. Het belangrijkste thema in dit proefschrift is het toetsen van de geldigheid van enkele, in de literatuur gesuggereerde, redenen voor het beleggen via beleggingsfondsen. Een reden die veel aandacht krijgt in de literatuur is de bewering dat fondsmanagers beschikken over speciale kennis of deskundigheid in het selecteren van aandelen zodanig dat het beleggingsfonds een interessant beleggingsprodukt wordt met rendementskenmerken die niet bereikbaar waren voor de individuele beleggers zelf. De performance van beleggingsfondsen kan worden geëvalueerd door te toetsen of beleggers mean-variance efficiënte set van hun huidige portefeuille kunnen uitbreiden door een positie te nemen in een beleggingsfonds. Dat wil zeggen of door het beleggen in het beleggingsfonds een hoger gemiddeld rendement kan worden behaald bij gelijkblijvend risico. Aan de hand van een literatuur overzicht worden in Hoofdstuk 2 de belangrijkste concepten van dit proefschrift geïntroduceerd. Er wordt aangetoond dat, onder standaard aannames, de hypothese dat een belegger zijn investeringsset niet kan uitbreiden, eenvoudig getoetst kan worden met behulp van regressie. Een alternatieve interpretatie van deze testen is het toetsen op outperformance van de beleggingsfondsen ten opzichte van een aantal referentie-activa, waarbij er impliciet wordt aangenomen dat de belegger momenteel al een efficiënte combinatie van de referentie-activa aanhoudt die corresponderen met een bepaald prijsvormingsmodel. Onder bepaalde aannames kan een vergelijkbare regressieanalyse worden gebruikt om te toetsen op outperformance conditioneel op economische omstandigheden.

De hypothese van persistentie in performance stelt dat beleggingsfondsen die in het verleden goed hebben gepresteerd ook goed blijven presteren. Toetsen van deze hypothese zijn gevoelig voor de keuze van de 'factor mimicking' portefeuilles van het onderliggende evaluatie model. Correctie van de performance voor publiek beschikbare informatie, zoals bijvoorbeeld een momentum factor, kan veel van de waargenomen persistentie in de rendementen van beleggingsfondsen verklaren. Een alternatieve interpretatie zou kunnen zijn dat beleggers die niet beleggen op basis van een momentum strategie, de investeringsset kunnen uitbreiden door een positie te nemen in beleggingsfondsen die actief een momentum strategie volgen. Tenslotte worden de gevolgen van zogenaamde overlevingseffecten (survival biases) besproken bij het gebruik van traditionele technieken voor het evalueren van de performance van beleggingsfondsen en de persistentie in deze performance, en worden correctiemethoden voor deze effecten voorgesteld.

In Hoofdstuk 3 worden de prestaties van internationaal beleggende Amerikaanse beleggingsfondsen geanalyseerd, waarbij wordt gecorrigeerd voor marktfricties zoals beperkingen op short posities en transactiekosten. Wanneer marktfricties afwezig zijn, laten we zien dat een risico-averse mean-variance optimaliserende Amerikaanse belegger zijn efficiënte investeringsset kan uitbreiden door te beleggen in internationaal beleggende fondsen, gegeven dat de belegger initieel een internationaal gediversificeerde portefeuille aanhoudt met gewichten die corresponderen met de marktkapitalisatie van de individuele landen. Voor sommige van de fondsen is dit resultaat robuust voor het meenemen van beperkingen op short posities en transactiekosten. Echter, voor beleggers die momenteel efficiënt in de Amerikaanse, Europese en Japanse aandelen indices beleggen, kan de hypothese dat beleggen in deze fondsen niet leidt tot uitbreiding van de efficiënte set niet worden verworpen wanneer er rekening wordt gehouden met restricties op short posities op de beleggingsfondsen als wel op de referentie-activa. Tenslotte laten we zien dat marktfricties eenvoudig kunnen worden meegenomen in toetsen op conditionele outperformance, waarbij wordt geconditioneerd op een verzameling informatie variabelen die verondersteld worden de toestand van de economie weer te geven. Het blijkt dat in sommige gevallen en bij bepaalde economische omstandigheden de internationaal beleggende fondsen ook in geval van fricties een uitbreiding geven van de efficiënte investeringsset van beleggers met een vooraf bepaalde landen weging.

Een eenvoudige manier om de performance te evalueren is om het gerealiseerde rendement van een fonds te vergelijken met het rendement op een referentie-actief dat overeenkomt met de beleggingsstijl van het beleggingsfonds. Vaak wordt er in deze zogenaamde relatieve performance meting een index gebruikt die overeenkomt met de door het fonds zelf gerapporteerde beleggingsstijl. In Hoofdstuk 4 wordt aangetoond dat stijl-analyse, gebaseerd op rendementen, gebruikt kan worden om de relatieve performance meting te verbeteren. Ten eerste, kan stijl-analyse gebruikt worden om op objectieve wijze de beleggingsstijl van het fonds te bepalen, en daarmee sterk samenhangend, kan stijl-analyse worden gebruikt als een instrument in relatieve performancemeting. Een zowel op relatieve als op risico gecorrigeerde performance evaluatie van een steekproef Nederlandse beleggingsfondsen laat zien dat de cash positie van de fondsen een belangrijk effect heeft op de relatieve performance van beleggingsfondsen. Verder wordt er aangetoond dat onder welke aanname relatieve performancemeting overeenkomt met performancemeting op basis van Jensen [1968]. Het meest opvallende resultaat is dat fondsen die voornamelijk in Nederlandse aandelen beleggen, een uitbreiding bieden van de mean-variance efficiënte set van Nederlandse beleggers die momenteel een internationaal gediversificeerde portefeuille aanhouden, zelfs na het opleggen van beperkingen op short posities.

In Hoofdstuk 5 worden vertekeningen geanalyseerd die zich in kleine steekproeven kunnen voordoen bij in de literatuur voorgestelde schattingsmethoden voor het toetsen op voorspelbaarheid van rendementen op fondsen. Een Monte Carlo simulatie experiment laat zien dat deze onzuiverheden in het algemeen niet verwaarloosbaar zijn. De vertekening correspondeert met een negatieve persistentie en kan dus geen 'hot hands' fenomeen genereren. In het hoofdstuk worden twee benaderingen voorgesteld die de onzuiverheid kunnen elimineren. De eerste methode vereist dat verwachte rendementen worden geschat over toekomstige waarnemingen, wat impliceert dat de meest recente waarnemingen niet worden gebruikt bij het schatten van korte termijn persistentie coëfficiënten. De tweede methode is gebaseerd op het gebruik van instrumentele variabelen bij het schatten van het model in eerste verschillen. Alhoewel deze methode consistente schattingen geeft, laat het Monte Carlo experiment zien dat de methode leidt tot nogal inefficiënte schattingen. Een empirische analyse suggereert de aanwezigheid van een klein persistentie effect in beleggingsfondsen die hoofdzakelijk in Amerikaanse groeiaandelen beleggen. In het bijzonder, suggereren de resultaten dat het ieder kwartaal selecteren van fondsen met een hoog rendement over de laatste vier kwartalen, ten opzichte van andere fondsen, het verwachte rendement op de portefeuille van beleggingsfondsen significant verhoogt.

Tenslotte is er in Hoofdstuk 6 geanalyseerd in hoeverre performance-evaluatie studies van beleggingsfondsen en studies van performance-persistentie beïnvloed worden door overlevingseffecten en wordt aangegeven hoe hier eventueel voor gecorrigeerd kan worden. Gebruikmakend van een zogenaamde overlevingseffect-vrije steekproef, is aangetoond dat er toch vertekeningen ontstaan indien de prestaties van een fonds met behulp van standaard technieken worden geëvalueerd. Om te corrigeren voor deze vertekeningseffecten is kennis van het overlevingsproces van essentieel belang. Daarom wordt er in dit hoofdstuk een proces gemodelleerd dat het overleven van fondsen beschrijft. Er wordt aangetoond dat de in het verleden behaalde rendementen, de leeftijd van het fonds en algemene macro-economische schokken een significant effect hebben op de overlevingskansen van beleggingsfondsen. Een Monte Carlo simulatie experiment laat zien dat het overlevingsproces leidt tot een schijnbaar persistentie patroon in de performance. Verder is er aangetoond dat het overlevingsmodel gebruikt kan worden om te corrigeren voor overlevingsvertekeningen in performance- meetmethoden. De correctiemethode wordt gebruikt om de persistentie te schatten in de performance van Amerikaanse groei, aggressieve groei en inkomensfondsen. Gebruikmakend van een vier-factor model, met factor mimicking portefeuilles voor grootte, boekwaarde ten opzichte van marktwaarde, momentum en een marktportefeuille, is aangetoond dat er geen bewijs is voor de bewering dat beleggingsfondsen die in het verleden goed hebben gepresteerd na correctie voor deze factoren, deze prestatie blijven voortzetten.

Center for Economic Research, Tilburg University, The Netherlands
Dissertation Series

No.	Author	Title
1	P.J.J. Herings	Static and Dynamic Aspects of General Disequilibrium Theory; ISBN 90 5668 001 3
2*	Erwin van der Krabben	Urban Dynamics: A Real Estate Perspective - An institutional analysis of the production of the built environment; ISBN 90 5170 390 2
3	Arjan Lejour	Integrating or Desintegrating Welfare States? - a qualitative study to the consequences of economic integration on social insurance; ISBN 90 5668 003 x
4	Bas J.M. Werker	Statistical Methods in Financial Econometrics; ISBN 90 5668 002 1
5	Rudy Douven	Policy Coordination and Convergence in the EU; ISBN 90 5668 004 8
6	Arie J.T.M. Weeren	Coordination in Hierarchical Control; ISBN 90 5668 006 4
7	Herbert Hamers	Sequencing and Delivery Situations: a Game Theoretic Approach; ISBN 90 5668 005 6
8	Annemarie ter Veer	Strategic Decision Making in Politics; ISBN 90 5668 007 2
9	Zaifu Yang	Simplicial Fixed Point Algorithms and Applications; ISBN 90 5668 008 0
10	William Verkooijen	Neural Networks in Economic Modelling - An Empirical Study; ISBN 90 5668 010 2
11	Henny Romijn	Acquisition of Technological Capability in Small Firms in Developing Countries; ISBN 90 5668 009 9
12	W.B. van den Hout	The Power-Series Algorithm - A Numerical Approach to Markov Processes; ISBN 90 5668 011 0
13	Paul W.J. de Bijl	Essays in Industrial Organization and Management Strategy; ISBN 90 5668 012 9

---

* Copies can be ordered from ThelaThesis,Prinseneiland 305, 1013 LP Amsterdam, The Netherlands, phone + 31 20 6255429; fax: +31 20 6203395; e-mail: office@thelathesis.nl

No.	Author	Title
14	Martijn van de Ven	Intergenerational Redistribution in Representative Democracies; ISBN 90 5668 013 7
15	Eline van der Heijden	Altruism, Fairness and Public Pensions: An Investigation of Survey and Experimental Data; ISBN 90 5668 014 5
16	H.M. Webers	Competition in Spatial Location Models; ISBN 90 5668 015 3
17	Jan Bouckaert	Essays in Competition with Product Differentiation and Bargaining in Markets; ISBN 90 5668 016 1
18	Zafar Iqbal	Three-Gap Analysis of Structural Adjustment in Pakistan; ISBN 90 5668 017 x
19	Jimmy Miller	A Treatise on Labour: A Matching-Model Analysis of Labour-Market Programmes; ISBN 90 5668 018 8
20	Edwin van Dam	Graphs with Few Eigenvalues - An interplay between combinatorics and algebra; ISBN 90 5668 019 6
21	Henk Oosterhout	Takeover Barriers: the good, the bad, and the ugly; ISBN 90 5668 020 x
22	Jan Lemmen	Financial Integration in the European Union: Measurement and Determination; ISBN 90 5668 021 8
23	Chris van Raalte	Market Formation and Market Selection; ISBN 90 5668 022 6
24	Bas van Aarle	Essays on Monetary and Fiscal Policy Interaction: Applications to EMU and Eastern Europe; ISBN 90 5668 023 4
25	Francis Y. Kumah	Common Stochastic Trends and Policy Shocks in the Open Economy:Empirical Essays in International Finance and Monetary Policy; ISBN 90 5668 024 2
26	Erik Canton	Economic Growth and Business Cycles; ISBN 90 5668 025 0
27	Jeroen Hoppenbrouwers	Conceptual Modeling and the Lexicon; ISBN 90 5668 027 7

No.	Author	Title
28	Paul Smit	Numerical Analysis of Eigenvalue Algorithms Based on Subspace Iterations; ISBN 90 5668 026 9
29	Uri Gneezy	Essays in Behavioral Economics; ISBN 90 5668 028 5
30	Erwin Charlier	Limited Dependent Variable Models for Panel Data; ISBN 90 5668 029 3
31	Rob Euwals	Empirical Studies on Individual Labour Market Behaviour; ISBN 90 5668 030 7
32	Anurag N. Banerjee	The Sensitivity of Estimates, Inferences, and Forecasts of Linear Models; ISBN 90 5668 031 5
33	Frans A. de Roon	Essays on Testing for Spanning and on Modeling Futures Risk Premia; ISBN 90 5668 032 3
34	Xiangzhu Han	Product Differentiation, Collusion and Standardization; ISBN 90 5668 033 1
35	Marcel Das	On Income Expectations and Other Subjective Data: A Micro-Econometric Analysis; ISBN 90 5668 034 X
36	Jeroen Suijs	Cooperative Decision Making in a Stochastic Environment; ISBN 90 5668 035 8
37	Talitha Feenstra	Environmental Policy Instruments and International Rivalry: A Dynamic Analysis; ISBN 90 5668 036 6
38	Jan Bouwens	The Use of Management Accounting Systems in Functionally Differentiated Organizations; ISBN 90 5668 037 4
39	Stefan Hochguertel	Households' Portfolio Choices; ISBN 90 5668 038 2
40	Henk van Houtum	The Development of Cross-Border Economic Relations; ISBN 90 5668 039 0
41	Jorg Jansen	Service and Inventory Models subject to a Delay-Limit; ISBN 90 5668 040 4
42	F.B.S.L.P. Janssen	Inventory Management Systems: control and information issues; ISBN 90 5668 041 2
43	Henri L.F. de Groot	Economic Growth, Sectoral Structure and Unemployment; ISBN 90 5668 042 0

No.	Author	Title
44	Jenke R. ter Horst	Longitudinal Analysis of Mutual Fund Performance; ISBN 90 5668 043 9

JENKE TER HORST
Groningen in 1993. Since October 1994, he has been a Ph.D. student at Tilburg University. His main interest is in financial econometrics. Since August 1998, he is affiliated to the Department of Finance of Tilburg University.

Many different motivations for investing in mutual funds have been provided in the literature, including the claim that managers of mutual funds have special abilities that can be used to outperform the market. Testing of the validity of these claims is complicated by two facts. First, the expected returns on mutual funds show cross-sectional as well as time series variation. Second, mutual funds that did not do very well in the past tend to stop trading more often than other funds. The latter implies that an analysis of returns on mutual funds that are currently traded is possibly affected by so-called survivorship bias. The aim of this thesis is to use longitudinal econometric techniques to test the validity of some of the motivations for investing in mutual funds that have been given in the literature.

ISBN 90 5668 043 9

CentER Dissertation